Ghent University
Faculty of Sciences, Department of Biology
Phycology Research Group

Gamete differentiation: a genetic, biochemical and evolutionary study using *Ectocarpus siliculosus*

by

Agnieszka Lipinska

Promotor: Prof. Dr. Olivier De Clerck
Co-promotor: Prof. Dr. Els Van Damme

Academic year 2013-2014

This thesis is submitted as a partial fulfillment of the requirements for the degree of Doctor (PhD) in Science (Biology)

This thesis is dedicated to my husband, Mateusz
and to my loving parents.

The research reported in this thesis was funded with a PhD BOF Grant 09/24J/117, Ghent University, Belgium and performed in the Research Group Phycology and Centre for Molecular Phylogenetics and Evolution at Ghent University.

Examination committee

Promotor

Prof. Dr. Olivier De Clerck
Ghent University

Co-promotor

Prof. Dr. Els Van Damme
Ghent University

Members of the reading committee

Dr. Mark Cock
Station Biologique de Roscoff

Dr. Sofie Derycke
Ghent University

Prof. Dr. Koen Sabbe
Ghent University

Other members of the examination committee

Dr. Lieven Sterck
VIB – Ghent University

Prof. Dr. Annemieke Verbeken
Ghent University

Acknowledgements

I would like to thank my supervisors Olivier De Clerck and Els Van Damme for giving me the opportunity to work on this thesis. I owe you a huge debt of gratitude for being so unflinchingly positive, understanding and encouraging! You have both ensured I have stayed focused and motivated while doing a PhD. Special "Thank you" goes to Olivier, upon whom the burden of my supervision has fallen most heavily; he has proved to be a fantastic, friendly and wise supervisor.

A number of people have made invaluable contributions, without which this thesis would not have been possible. Particular thanks go to Dieter Müller, whose original work formed the inspiration for this project, who guided me through the intricate field of algal culturing and together with Akira Peters kindly provided *Ectocarpus* strains which have been essential to this study. Erwan Corre, Susana Coelho and Mark Cock from Biological Station in Roscoff also deserve a special mention for helpful discussions and suggestions during my first steps in Next Generation Sequencing data analysis. Special thanks go to Sofie D'hondt who thought me all about molecular lab techniques, but also much more. You have been there as a friend and confident throughout my time in Gent and it would not have been the same without you!

I thank all my fellow lab members and visitors, especially Frederik L., Heroen, Ana, Lennert, Kenny, Frederique, Viviana, Christophe, Frederic M. and Marie, for your energy, entertaining talks and general gossiping during lunch breaks, and the crazy Christmas parties! I thank Christelle and Anja, for all the help with administrative work and for your contagious optimism. Many thanks to the members of Protistology and Aquatic Ecology group, especially Jeroen, Evelien, Pieter, Lander, Celine, Ilse, Renaat, Olga and Dagmar; you made the work environment very 'user friendly'!

The friends I have made during this time have also been central in making the bad times survivable and the good times so enjoyable. My special thanks are to my polish girls 'gang', especially to Paula, Ania, Agata, Kamila, Kasia's, Asia's and Agnieszka's; for your friendship, all the laughter and for making me feel at home. Dagmar, Evelien, Frederique and Mila, you gave a new dimension to needlework and made me adore our Thursday evenings. My thanks to Gio, Jeroen, Hugues and Nico for friendship and many precious memories along the way.

I cannot forget my dear friends in Poland, who stayed by my side and supported me despite the distance. Ewa and Magda you kept me smiling and helped keep things in perspective.

Of course no acknowledgments would be complete without giving thanks to my family, in particular my parents, my grandparents, my aunt, my brother and my little 'sisters' Ada and Nicol who have been so supportive and encouraging throughout these years in Gent, and beyond. They have always expressed how proud they are of me and how much they love me.

Last, but certainly not least, my tremendous and deep thanks to my best friend and husband, Mateusz. Through his love, patience, support and unwavering belief in me, I have been able to complete this long dissertation journey.

Thank you all!
Dziękuję wam wszystkim!

Table of content

Chapter 1

General introduction and thesis outline

'Sex is the queen of problems in evolutionary biology. Perhaps no other natural phenomenon has aroused so much interest; certainly none has sowed as much confusion. The insights of Darwin and Mendel, which have illuminated so many mysteries, have so far failed to shed more than a dim and wavering light on the central mystery of sexuality, emphasizing its obscurity by its very isolation.'

/Graham Bell in "The Masterpiece of Nature: The Evolution of Genetics and Sexuality"/

1 The evolution of sexual reproduction

The prevalence of sexual reproduction is so wide among eukaryotes that its benefits and usefulness seem to be unquestionable. Sex is present everywhere in nature and it affects almost every aspect of life, including morphology and behavior. It is then surprising to many people that the evolution of sex raises so much controversy.

Sex as a mechanism of transfer and recombination of genetic material exists in eukaryotic as well as prokaryotic organisms (Maynard Smith, 1978b)(Levin, 1988). However, certain aspects of sex like the cellular mechanisms, the frequency and local rather than global recombination differ substantially in prokaryotes when compared to eukaryotic systems (Maynard Smith, 1990)(Maynard Smith *et al.*, 1991). Sex in bacteria and archaebacteria involves unidirectional inheritance of DNA from another source than the parental cell through for example conjugation or transduction. Among prokaryotes this process is independent of gamete fusion or reproduction and creates variation through recombination of existing genes or introduction of novel sequences (e.g. plasmids). In eukaryotic sex, the process of DNA recombination is accompanied by a succession of meiosis and syngamy (Maynard Smith, 1978b) and as such is discussed further in this chapter.

Current theory suggests that the last common ancestor of all eukaryotes was a facultative sexual unicellular species, implicating that sex has a single evolutionary origin (Dacks & Roger, 1999). Despite considerable variations in the sexual cycle, coupled to ploidy changes and modes of reproduction, the core elements are remarkably conserved across all lineages of eukaryotes, from algae over slime mould to humans.

When talking about the evolution of sexual reproduction, we focus on two main questions: the origin of sexual reproduction and its maintenance. Both of these aspects are difficult to

study experimentally; both are subjects of vivid scientific dispute. The following paragraphs aim to summarize the variety of current, sometimes competing, hypotheses about the evolution and nature of sex.

1.1 The origin of sex

According to Cavalier-Smith (2002) the onset of sexual reproduction can be linked to neomuran revolution, during which the common ancestor of eukaryotes and archaebacteria evolved histones for stabilization of chromosomes and accordingly new DNA-handling enzymes. The substitution of the bacterial peptidoglycan cell wall by glycoproteins and the evolution of nuclei, cytoskeleton, cell cycle control and mitosis in eukaryotes allowed them to undergo cell fusion, a fundamental prerequisite for sex (Cavalier-Smith, 2002).

Successful sexual reproduction in eukaryotes requires cell fusion (syngamy) followed by nuclear fusion (karyogamy) subsequent to meiosis, resulting in the alternation of haploid and diploid generations (Bell, 1982)(Maynard Smith & Szathmáry, 1998). However the transition from haploid to diploid life history phases might be older than sex itself and can result from spontaneous diploidization by endomitosis (Cleveland, 1947)(Hurst & Nurse, 1991). Meiosis would then have evolved as a result of selection for periodic ploidy reduction for ecological reasons, providing that different phases are more suited for specific environments (Margulis & Sagan, 1990).

Sex as a DNA repair mechanism

Diploid organisms hold also another big advantage – they can repair double-strand damaged DNA by homologous recombination using a sister chromatid (available in the G2 phase after DNA replication) or a homologous chromosome as a template. So temporary diploidy could be a response to high concentrations of free radicals damaging DNA (Szathmary *et al.*, 1990). In fact, the advantage of recombinational repair and genetic complementation is one of the leading theories on the origin of sex (Dougherty, 1955)(Bernstein *et al.*, 1985)(Bernstein *et al.*, 1987).

Whereas, in the beginning syngamy would not have been necessary for diploidization, nor was crossing-over a feature of a haploid-diploid cycle (Hurst & Nurse, 1991), the benefits of hybrids later preponderated for syngamy (Maynard Smith & Szathmáry, 1998). The obvious example of such a benefit is the masking of deleterious recessive mutations in a heterozygous hybrid, resulting in a higher fitness of the individual.

Although in many lineages sex is obligatory and connected directly to reproduction (the latter cannot occur without the former), in many protists sex occurs sporadically in response to unfavorable conditions. An interesting example is provided by the facultative sexual green

alga *Volvox carteri,* in which an increase of reactive oxygen species (ROX) concentration triggers the expression of sex cycle genes (Nedelcu *et al.*, 2004). This implicates that sexual reproduction could be an adaptive defense mechanism against DNA damage due to oxidative stress.

Muller's ratchet

An additional explanation to account for the success of sexual reproduction is given by "Muller's ratchet". Herman Joseph Muller predicted that deleterious mutations would accumulate over generations in the genomes of asexual species and eventually would contribute to their extermination, whereas sexual reproduction is capable of producing 'mutation-free' individuals through recombination (see Box 1) (Muller, 1964). Moreover, recombination generates genetic variability, which allows species to adapt to novel environmental conditions and withstand competition, assuming that their effective population sizes are large enough (Colegrave, 2002)(Colegrave, 2012). This is further supported by the fact that asexual lineages are in general occupying the terminal nodes of the tree of life, implicating they have a short evolutionary lifespan and high extinction rate (Maynard Smith, 1978b). As a result of recombination, a sexual population will on average have a lower number of mutations and a higher fitness than an asexual population. However, the selection favoring modifiers of sex and recombination will overcome sex related costs only in case when mutation rates are high (Charlesworth, 1990).

Sex and the selfish gene

The spread of modifiers of sex relates to another theory on the very origin of sexual reproduction, which refers to propagation of selfish genetic elements. These elements (i.e. transposon elements, genes, chromosomes) possess characteristics that enhance their own spread relative to the rest of an individual's genome (Dawkins, 1989). Transposable or selfish genes that carry information specifying sexual reproduction would facilitate their own transmission, because selfish genes cannot spread within an asexual population with a strictly clonal reproduction (Cavalier-Smith, 1980).

1.2 The maintenance of sex

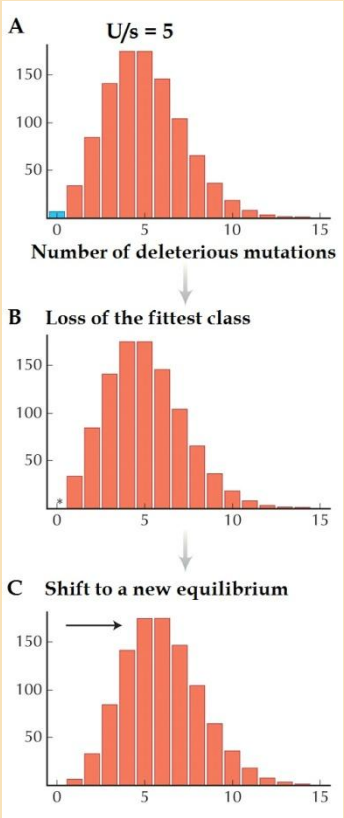
The second important aspect of evolution of sexual reproduction is its maintenance. The benefit of sex in relation to its cost is a matter of debate since Darwin's "*The Descent of Man and Selection in Relation to Sex*" (Darwin, 1872). It is clear, that despite the hindrances, sexual reproduction must present exceeding advantages, because almost 99.9% of eukaryotes engage in sex (Vrijenhoek, 1998). However, drawing well-defined boundaries separating

strengths from weaknesses and deciding which of the various possibilities actually play a role in sex maintenance is difficult. Genetic properties, lifestyles and environmental conditions could influence the net value of sex which also relies upon the adopted reference point. Namely, the benefits of sex will be different if we look from a perspective of a gene, an individual or a group of individuals. The following paragraphs present a condensed overview of these tangled attributes of sexual reproduction.

Box 1 | Muller's ratchet - accumulation of deleterious mutations in asexual populations

Harmful as well as beneficial mutations occur spontaneously in all organisms. However, as it was pointed out by H.J. Muller (1964): *"... an asexual population incorporates a kind of ratchet mechanism, such that it can never contain in any of its lines, a load of mutations smaller than that already existing in its present least-loaded lines...".*

This means that without occasional recombination, an asexual lineage can never create a line free of harmful mutations, which can eventually drive it to extinction. Likewise, without recombination it is not possible to combine favorable mutations in a single line. The accumulation of deleterious mutations in asexual species is represented on the chart:
(**A**) A balance between mutation (U=0.1) and selection (s=0.02) rate leads to equilibrium of, on average, U/s=5 deleterious mutations per genome. However, only $1000 \cdot e^{-U/s}=6.7$ individuals are mutation-free in a population of 1000 genomes (blue bar).
(**B**) This fittest class will eventually be lost by chance. Without recombination or back mutation, the mutation-free individuals cannot be recovered.
(**C**) The whole distribution shifts to the right in one click of Muller's ratchet and this process is irreversible.



(Adapted from (Barton et al. 2007))

1.2.1 The advantages of sex

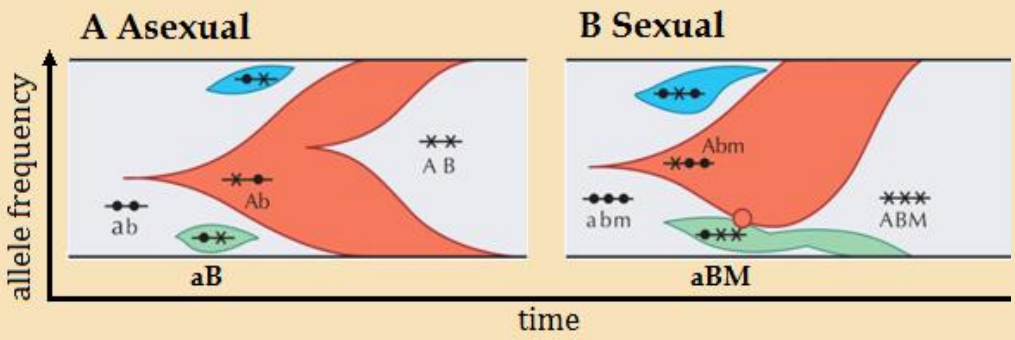
Sex as a generation of variation

Sex as a generation of variation on which the selection can act is one of the most popular hypotheses proposed by Weismann et al. (1889) and later elaborated by Fisher (1930) and

Muller (1932). In this theory, sexual reproduction allows beneficial mutations that come from different individuals to be combined in the offspring (Fisher-Muller hypothesis) (see Box 2). As a result, in the long-term, recombination facilitates fixation of beneficial mutations and/or separation of profitable genes from disadvantageous mutants, and thus enhances the potential for evolution, because recombination expands the range of gene combinations exposed to natural selection (Williams, 1975)(Maynard Smith, 1978b) (Bell, 1982) (Crow, 1988) (Barton, 1995)(Goddard *et al.*, 2005)(Barton *et al.*, 2007).

Natural selection can act by producing variable offspring which carry in some cases more deleterious mutations and in some case fewer mutations then its ancestors. As the fittest genome spreads by selection, so are the genes controlling the rate of sex and recombination (so called 'modifiers') (Charlesworth, 1990). Ronald Fisher also suggested that sex might untangle the advantageous genes from their genetic surrounding (i.e. deleterious genes) and facilitate their spread, thus improving the response to selection and elimination of harmful alleles (Fisher, 1930)(Otto, 2009).

Box 2 | Fisher-Muller hypothesis – sex as a generation of genetic variation



(**A**) Beneficial mutations in asexual populations must be established sequentially. As an example, if allele A is to replace allele a in the population, then any other favorable allele that occurs at other loci (e.g. allele B) can be fixed only if it appears in the same genome as allele A. Different colours represent genotype frequencies in a population over time.
(**B**) In sexual populations, beneficial alleles can be placed together in one genome through recombination. This leads to an advantage of modifiers that control sex and recombination. As an example favorable allele B is joined with favorable allele A by recombination (red circle). A modifier allele M, which is required in the process, increases in frequency by hitchhiking.

(Adapted from (Barton et al. 2007); redrawn from (Barton & Charlesworth 1998)).

Sex releases the speed limit of evolution

Studies on *Chlamydomonas* showed that sex increases the rate of adaptation and fitness, but depending on the population size (Colegrave, 2002). Larger populations experience substantial benefits from sex, whereas a small population almost did not differ from the asexual controls. The results indicate that a limited number of mutations encountered in small populations is not a satisfactory selection of variants for adaptation, in contrast to large populations, where mutations are plentiful, and sex can release the speed limit of evolutionary adaptation (Colegrave, 2002).

The Red Queen hypothesis

Although it seems that gene shuffling may be unfavorable because it can break beneficial gene combinations of parents, rapid environmental changes experienced by progeny could favor previously neutral or deleterious alleles. Such rapid changes in the environment can be caused by co-evolution between hosts and parasites, and recombination can aid sexual individuals to escape infection (Hamilton *et al.*, 1990)(Peters & Lively, 1999)(Salathé *et al.*, 2008). This theory, known as the Red Queen Hypothesis, is widely discussed as a support for persistence of sex (Bell, 1982).

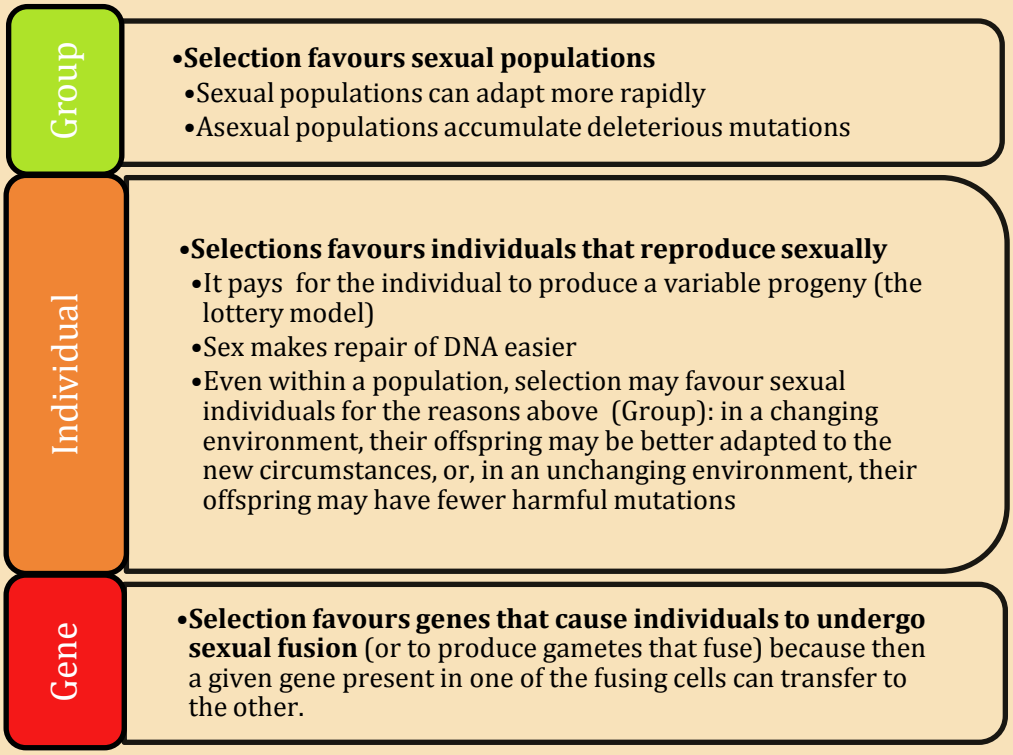
The Red Queen hypothesis could be observed in course while studying sexual and asexual populations of the snail *Potamopyrgus antipodarum* (Jokela *et al.*, 2009). As asexual female snail became more spread, they encountered higher rates of parasite infections causing the most common clones to be replaced by rare clones and the entire population to decrease dramatically in number within a few years of observation. Sexually reproducing snails, however, persisted showing much more stability over time. Additionally, the common clones, which were initially more resistant, became more prone to infection by sympatric parasites over the same period of time (Jokela *et al.*, 2009).

Molecular advantages of recombination

Two molecular advantages can be attributed to sex. As mentioned earlier, diploid organisms can deal with DNA damage through recombinational repair with homologous chromosomes, which is argued to be the main drive for recombination (Bernstein *et al.*, 1988)(Michod & Levin, 1988)(Cox, 1993). Indeed, many proteins that are involved in recombination are necessary for DNA repair (i.e. the RecA, RecE, RecJ, and RecQ proteins in *Escherichia coli* (Cox, 1993)(Kusano *et al.*, 1994) and their homologs RAD51, RAD52, and RAD54 proteins in *Saccharomyces cerevisiae* (Basile *et al.*, 1992)). Mutants of these genes often fail to recombine and replicate in face of DNA damage.

The second molecular advantage of outcrossing is masking of deleterious alleles, which depends on heterozygosity of individuals. Sexual populations have the ability to avoid Muller's ratchet by producing offspring with a variable number of mutations compared to parents (Muller, 1964) and escape the effects of recessive deleterious mutations through dominance (Chasnov, 2000) or epistasis (Kimura & Maruyama, 1966)(Kondrashov, 1982)(Kondrashov, 1988)(Charlesworth, 1990). Furthermore, genotypes combining a large number of mutations and suffering low fitness through negative epistasis will be removed from the population, resulting in an overall more fit population (Kondrashov, 1988). In this theory, known as the "deterministic mutation hypothesis" (Kondrashov, 1988), sex benefits the group by compartmentalization of deleterious mutations.

Box 3 | Benefits of sexual reproduction – theories on the group, individual and gene level



(From Maynard Smith & Szathmary 1999)

Many theoretical models explaining how the major population genetic processes could give an advantage to sexual reproduction have been introduced over the last three decades (Maynard Smith, 1978b)(Michod & Levin, 1988)(Barton & Charlesworth, 1998)(Colegrave, 2002). However, too little evidence is currently available from natural populations to

accurately justify which of the presented theories are most suitable to settle the paradox of sex. There is likely to be truth in each of them and it is probable that they all contributed, to some extent, to the evolution and maintenance of sexual reproduction.

1.2.2 The cost of sex

The problem with sex becomes evident, when investigating the influence of sex on an individual's fitness. Sexual reproduction can be costly and it may resemble a steeplechase. Engaging individuals have to invest in the production of gametes, find an appropriate partner, attract it, recognize it and fuse to form a new organism. Accompanying these direct costs is often competition for access to the opposite sex, which is a foundation for sexual selection, a concept introduced already by Darwin in "*On the origin of species*"(Darwin, 1859) and expanded by him in 1871 in "*The Descent of Man and Selection in Relation to Sex*". Pathogen infections, predator attack, sexually transmitted diseases are other obstacles presented to the reproducing entities. At the molecular level, sexual reproduction can break apart beneficial gene combinations creating a new network that may decrease the survival rate of the progeny. Additionally, there are costs associated with the mode of reproduction. In particular, isogamous species need to deal with recombination, cellular-mechanical costs, fertilization and genome dilution whereas additional costs associated with anisogamy are production of males and sexual selection (Lewis, 1987).

Two-fold cost of sex

In 1971 John Maynard Smith redefined the costs and benefits, and described the "two-fold cost of sex", by the observation that most of the organisms produce two types of gametes – a large egg and a small sperm. In consequence, females allocate much more resources to the offspring than males. Furthermore, if females invest half of their reproductive potential into production of males whose only role is to donate their gene pool to the zygote, asexual populations are twice as much efficient in the production of a progeny (Maynard Smith, 1978b) (hence the 'term two-fold cost of sex'). This is why the two-fold cost of sex is also recognizable as the cost of males. Considering there is no investment by males into the production of a progeny, a female asexual mutant cloning itself will produce four times more daughters compared to a sexual female after only two generations (assuming sex ratio 1:1; see Fig. 1).

However, the twofoldness is valid only if males never assist in the production of offspring and allocate all their reproductive resources into maximizing their own fertilization success. In case of paternal care the cost of sex owing to male will deviate from two, especially in a monogamous population where males evolve traits that maximize female's reproductive

success (Lehtonen *et al.*, 2012). Isogamous populations can also escape the two-fold cost if both parents invest only in gamete production; since there is no gamete size difference, the reproductive output can then be compared to the asexual mode.

However, the two-fold cost can also be increased if we encounter a sexual conflict, particularly when female reproductive success is decreased by male competitiveness. Such effect is observed in *Drosophila melanogaster*, where male seminal fluids elevate female egg-laying, annihilate sperm from other mates and reduce female receptivity to further matings, but also increase female death rate (Chapman *et al.*, 1995). Therefore the cost of sex in females is heightened by evolutionary conflicts between males. Other examples include the beetle *Callosobruchus maculatus,* whose male spiky genitalia generate damage to the female reproductive tract resulting in reduced female longevity (Eady *et al.*, 2007). The adaptive harm hypothesis suggests that a harmed female would have a reduced disposition to remate which benefits male individuals (Johnstone & Keller, 2000); alternatively punctures which give males a reproductive advantage (i.e. increased delivery of seminal fluids or anchoring during copulation) are a side effect of sperm competition (Parker, 1979).

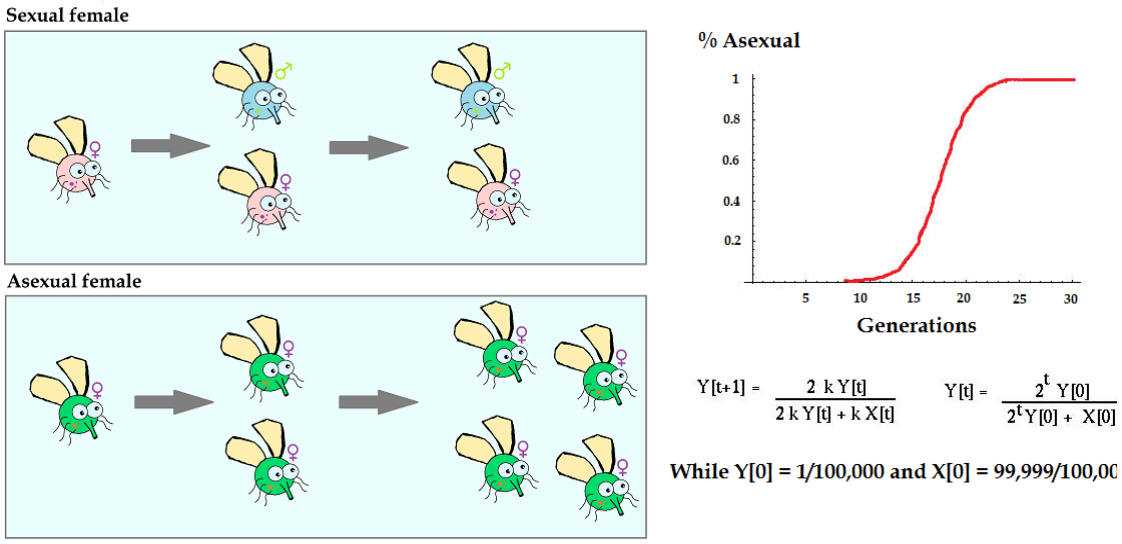


Figure 1. Spread of asexuality assuming a two-fold cost of sex. The hypothetical sexual and asexual populations of fly are shown with the offspring of two subsequent generations. The asexual population, which does not invest in males, can double in frequency over time. The frequency of asexuals over generations is shown on the graph, assuming sexuals producing k offspring and asexuals producing 2k in one generation. X – frequency of sexuals, Y – frequency of asexuals.

Genome dilution

Another theory that invokes a two-fold cost of sex, but owned it to the genome dilution, was recognized by George Williams and is known as the "cost of meiosis" (Williams, 1975). Genome dilution refers to the fact that an asexual female passes 100% of the genetic material to its offspring, whereas a sexual female transmits only 50% of her material. In other words, sexual females are only half related to their offspring in contrast to the asexual ones. However, it was pointed out that this phenomenon is meaningful only for genes that control sexual reproduction, diluting the rest of the genome is irrelevant (Treisman & Dawkins, 1976)(Barash, 1976)(Lehtonen *et al.*, 2012). These sex genes need to be passed on to the offspring to maintain the reproductive mode. Because two individuals engaging in sex should by definition possess genes determining sexual reproduction, a scenario in which a sexual population was invaded by an asexual population and genes coding for sex would be replaced by asexual ones, is highly unlikely (Colegrave, 2012). Additionally, the genes coding for sex 'recognize' each other and 'reunite' in the offspring, because they are the cause of mating in the first place, which negates the genome dilution hypothesis (Lehtonen *et al.*, 2012). However, an exception could be provided by hermaphrodites and when asexual organisms still produce males (Charlesworth, 1980)(Uyenoyama, 1984)(Engelstädter, 2008).

Cost of meiosis

Meiosis itself also presents a cost, since in many unicellular species the time consumed by meiosis is 5-100 times greater than for mitosis (Lewis, 1983)(Lewis, 1987). Although selection acts to reduce the time of meiosis, the key mechanisms of the process like chromosome pairing cannot be accelerated. This time cost of sex is especially relevant for unicellular organisms, because meiosis and gamete production cannot occur in parallel with growth and development (Lehtonen *et al.*, 2012).

Mate search

The difficulty of finding mate partners is not a trivial problem in sexual reproduction, which is reflected by a large number of hermaphroditic species. Eppley and Jesson showed that adult mate-search efficiency is correlated with evolution of hermaphroditism in multicellular organisms and is one of the factors responsible for the wide range of breeding system variations observed in many taxa, such as plants (Eppley & Jesson, 2008). In case of low mate-search efficiency which is prevalent especially in angiosperms and marine free spawners, we can expect strong selection for increased recognition efficiency or gamete dispersal distance (Eppley & Jesson, 2008). The latter group can provide valuable insight into breeding system evolution, because it hosts numerous closely related species with different breeding modes.

Mechanistics of recombination

Finally, the most general cost of sex is that of the molecular nature of recombination. Although gene shuffling can bring together beneficial alleles and separate them from harmful mutants which is favorable over the long-term course of evolution, in the short term it can break profitable gene combinations (Maynard Smith, 1978b)(Lewis, 1987) before it creates new beneficial associations. This cost is shared equally between isogamous, anisogamous, unicellular and multicellular species. It does not depend on sexual conflict or efficiency of mate-searching. It does, however, lean on the selection over time and space (Lehtonen *et al.*, 2012). Assuming that the environment does not change, the new generation with shuffled genotypes will be, on average, less fit then their parents (Williams, 1975)(Maynard Smith, 1978b)(Lewis, 1987)(Crow, 1988), because a mixture of genes from two parents has no guarantee of functioning as good as in the original set. This reduction of fitness is known as the "recombination load". However, in real-world evolution, we meet many more complexities including rapid environmental changes which favor selection. The old genetic combinations established by past selection are improved through sex and recombination and improve the fitness of offspring, turning the recombination load to an advantage (Salathé *et al.*, 2009)(Otto, 2009).

2 The evolution of sexes

The existence of two separate sexes is an added puzzle in the evolution of sexual reproduction. The benefits of combining genetic information from two parents have been discussed earlier, but why these parents have to be of opposite sex remains a question. Although nature knows examples of multiple mating types i.e. in fungi (Raper, 1966), in the majority of cases mating involves only two types, male and female. Darwin described the existence of a large female egg and a small size male sperm as one of the central mysteries of evolutionary biology. Unequally-sized gametes underlie the evolutionary processes of sex ratios and sex differences in behavior and morphology through sexual selection. Nonetheless, gamete dimorphism is not required for sexuality. Many lower eukaryotes produce gametes of equal size (isogametes) (Hoekstra, 1982) and it is believed that isogamy was most likely the ancestral condition, and that anisogamy developed independently in many lineages. Irrespective of gamete size, many animals and plants species display co-sexuality (Jarne, 1993), which allows them to produce both male and female reproductive cells. In fact, hermaphroditism is probably the common ancestral state in plants (Charlesworth *et al.*, 2005) and was suggested to be the ancestral state of all chordates as well by Darwin (Darwin, 1859). Animal genetic sex determining systems may, however, have evolved from environmental sex determination (Janzen & Krenz, 2005).

Box 4 | Sex can be expensive…

Sexual displays and rituals can be enormously expensive in terms of energy, time and resources. Male bowerbirds build sophisticated nests, decorated in flower petals, feathers and pebbles to make it more appealing for a female (on the left). Male emperor penguins take care of the egg after mating until it hatches, without moving or feeding for about two months (on the right).

Image retrieved from:http://4.bp.blogspot.com/6yPpF1p9zc8/UQsxSzipSyI/AAAAAAAAJ4/4rhQF3elvks/s1600/Satin+bowerbird.jpg

Image credit: BrynJ [http://www.flickr.com/photos/brynj/15642290/]

…and dangerous!

Seeking mates or even copulating can be dangerous for many species. Male tree frogs in Central America call to draw their partners (on the left). However, frog's calling attracts also carnivorous bats, which arrive to the feast. Praying mantises display sexual cannibalism (on the right). After copulation the hungry female feeds on its mate. It was also indicated that cannibalized males are able to copulate longer increasing the chance of fertilization.

Image retrieved from: http://biologypop.com/the-red-eyed-tree-frog-agalychnis-callidryas-info/

Image retrieved from: http://commons.wikimedia.org/wiki/File:Praying_Mantis_Sexual_Cannibalism_European-

Evolution of dioecy

Dioecy, distinct male and female organisms, could evolve in two scenarios through androdioecy or gynodioecy intermediate pathways. Gynodioecy assumes a male-sterile mutant entering the co-sexual population and selection acting on the hermaphroditic morph to decrease the allocation of resource to produce female eggs. Alternatively, in androdioecy, a

female-sterile mutant enters the population inducing inhibition of male function in the co-sexual morph. Both scenarios lead to establishment of separate male and female organisms through frequency-dependent selection (Delph, 2003), where the fitness of the hermaphrodite might increase by biased production of gametes of the opposite function to the ones produced by a unisexual mutant, assuming that the mutant is high in frequency (Avise & Nicholson, 2011). However as argued by Charlesworth and Charlesworth (Charlesworth & Charlesworth, 1978) a male or female mutant may invade a hermaphroditic population only if the self-fertilization causes severe inbreeding depression. On the other hand, benefits of sexual specialization, could counter the advantage of self-fertilization and lead to evolution of separate sexes (Charlesworth & Morgan, 1991)(Charlesworth & Charlesworth, 1999)(Meagher, 2007). Studies on angiosperms revealed both factors to be involved in the evolution of dioecy in plants (Costich & Meagher, 1992)(Dorken *et al.*, 2002)(Costich, 1995)(Bram, 2002)(Gleiser & Verdú, 2004)(Eppley & Pannell, 2007). Similarly, many factors could have played a role in the evolution of various breeding systems across other eukaryotic taxa (Hurst & Werren, 2001).

Sex determination

Sex determination takes its beginning when one member of a chromosome pair acquires a sex-determination function (Carvalho, 2002), however to produce genetically distinct male and female individuals two separate mutations have to occur (Charlesworth *et al.*, 2005). These loci are at the base of genetic control of sexual phenotypes and often are homozygous in one sex and heterozygous in the other (Bull, 1983). In addition or as an alternative system, one or several autosomal nuclear sex-determining loci, possibly influenced by cytoplasmic genes and environmental cues can be present (Korpelainen, 1998)(Ainsworth, 2000)(Janzen & Krenz, 2005). The complexity of this system ranges from single sex alleles to morphologically and genetically different chromosomes with arrested recombination (Bull, 1983) (Bachtrog *et al.*, 2011).

Despite the degree of evolutionary advancement, the sex chromosomes or sex determining regions show certain similarities; they evolved independently in many lineages and are distinguished by genetic degeneration due to the lack of recombination (Bull, 1983)(Charlesworth, 1996). The inhibition of recombination originated probably to suppress disadvantageous sexual phenotypes that would result from mixing sex-determining genes (Charlesworth, 1996). With time, sex-loci expanded their non-recombining regions through events like chromosome inversions, which contributed to the heteromorphic evolution of sex chromosomes by introducing other genes, not involved in sex determination, to the pool (Charlesworth & Charlesworth, 2005). The evolutionary pressure for a wider scope of

suppressed recombination applied to alleles that are sexually antagonistic, i.e. ones that are beneficial for male and harmful for female. Inclusion of such loci in the sex determining region presents clear advantages (Bull, 1983)(Rice, 1987). These mechanisms apply not only to the genes residing on the same chromosome as the sex locus, but also on the other autosomal chromosomes (Charlesworth & Charlesworth, 1980).

The complexity and variation in architecture of sex chromosomes suggest that they are products of independent evolution in different organisms (Fraser & Heitman, 2005)(Ming & Moore, 2007). Nevertheless, convergence of sex chromosomes structure regarding the heteromorphism in size and shape can often be observed among various systems (Ming & Moore, 2007). The most familiar heteromorphic sex-determination system is the XX/XY, as it is found in humans and most other mammals, as well as in some insects. Males are the heterogametic sex with two distinct chromosomes, whereas females carry two copies of chromosome X. In ZZ/ZW sex-determining systems, found in birds, some reptiles and some insects, the heterogametic sex (ZW) is a female characteristic and homogametic sex (ZZ) describes males. However, recent work has shown that the evolution of sex chromosomes is not dependent on a diploid genome (for a review see (Bachtrog *et al.*, 2011)). Dimorphic sexes can be genetically determined in haploid stages and are referred to as carrying U and V chromosomes. This system is best developed in bryophytes (Yamato *et al.*, 2007)(McDaniel *et al.*, 2007) and macroalgae (Cock *et al.*, 2010)(Blouin *et al.*, 2011), where females bear the U chromosome, male bear the V chromosome and the diploid sporophyte is always heterogametic (UV).

2.1 The evolution of mating types in isogamous populations

Isogametes are equal in size and morphologically identical. They can be distinguished by physiological differences, in which case they are often designated as "+" and "–" mating types. Studies on the evolution of mating types in isogamous populations are useful in understanding the forces shaping the evolution of distinct sexes, because it is widely accepted that sexual differentiation into two morphologically indistinguishable mating types has preceded the evolution of anisogamy.

Mating types and cytoplasmic elements

Theoretical analyses indicated that mating type evolution could be a consequence of selection for more efficient gamete recognition, and also a result of intragenomic conflict between nuclear and cytoplasmic DNA (Hoekstra, 1987). Indeed the two mating types show divergence in inheritance of mitochondrial or chloroplast genes (Birky, 2001). In *Chlamydomonas reinhardtii* chloroplast genomes are transmitted by the mating type plus

(*mt*+) parent and mitochondrial genomes by the mating type minus (*mt*-) parent (Gillham *et al.*, 1987). The conflict between cytoplasmic elements or selection for good competitors could affect the zygote fitness (Hurst & Hamilton, 1992)(Hutson & Law, 1993). Therefore nuclear mutations enforcing uniparental inheritance are favored by selection (Hoekstra, 1982)(Hurst & Hamilton, 1992)(Hutson & Law, 1993). Another explanation comes from the observations of evolutionary adaptations between mitochondrial and nuclear genomes (Blier, 2001)(Dowling *et al.*, 2008). Many mitochondrial genes migrated to the nucleus following the endosymbiotic event giving rise to mitochondria (Gray, 1999)(Esser, 2004). It is argued that that selection for mitonuclear co-adaptation may be driving uniparental inheritance in unicellular organisms, especially when large numbers of mitochondria are present (Hadjivasiliou *et al.*, 2012).

Mating types and pheromones

Other theoretical models predict that evolution of mating types would be greatly facilitated if a pheromone attraction system was already pre-existent, on which unipolarity of gamete recognition or adhesion could be superimposed (Hoekstra, 1982). In that model gamete A produces pheromone recognized by gamete B, but A and B are free to fuse with any gamete, so their mating behavior is still the same. Nevertheless, encounters between A and B are more common because of a chemotactic mechanism. Selection could subsequently act on other gamete characteristics in A and B. Given the genetic sex determination mechanisms, pheromone and recognition genes would be tightly clustered together on mating type loci determining various mating related functions (Hoekstra, 1982). In fact, DNA sequences of mating loci in fungi (i.e. the plant pathogen *Ustilago maydis* (Lee *et al.*, 1999) or the yeast *Cryptococcus neoformans* (Lengeler *et al.*, 2002)) revealed a single locus comprising homeodomain transcription factors, controlling developmental processes in sexual reproduction, pheromone producing genes and pheromone receptors.

2.2 The evolution of anisogamy

Sexual dimorphism is primarily reflected in anisogamy: larger eggs produced by the female and smaller sperm produced by the male. Anisogamy could be an example of how increasing levels of dimorphism may evolve after the establishment of two sexes, eventually leading to oogamy (large immotile eggs and small motile sperm). It is the most prevalent form of reproduction, which evolved several times independently among many lineages. Many theoretical models were developed to explain the success of anisogamy, linking it directly to gamete sizes and fitness (Parker *et al.*, 1972)(Maynard Smith, 1978b)(Bell,

1978)(Charlesworth, 1978)(Hoekstra *et al.*, 1984) (Matsuda & Abrams, 1999)(Dusenbery, 2000).

These models assume three selective forces:

1) selection for greater gamete productivity achieved by lowering the gamete size. So male sperm can be produced in larger quantity for the same cost;
2) selection for greater zygote size which have higher fitness, imposing selection on females to produce larger eggs for zygote viability;
3) selection for higher gamete encounter in external fertilization: a large immotile egg increases the chances of meeting a small motile sperm, since it offers a larger target surface.

An alternative approach suggests that small sperms prevent transmission of cytoplasmic symbionts, organelles and/or parasitic elements, enforcing uniparental organelle inheritance (Grun, 1976)(Cosmides & Tooby, 1981)(Hurst, 1990). Parasite mixing, due to cytoplasm mixing, at the time of zygote formation would create an additional cost of sex (Hurst, 1990), which could be the reason of substantial reduction of cytoplasm volume in the final stage of sperm maturation in several taxa (O'Donnell *et al.*, 2011).

3 Speciation and the genetic barriers to reproduction

One of the key problems in evolutionary biology is the origin of new species. Ever since Darwin it has been apparent that there is a strong relation between sex and speciation. As defined by Ernst Mayr – "Species are groups of actually or potentially interbreeding populations, which are reproductively isolated from other such groups" (Mayr, 1942), and further, each species has to evolve sufficient levels of discrimination to maintain its own identity (Mayr, 1964). This concept was elaborated by Dobzhansky, who formed a theory on the evolution of species, where biological isolation, that prevents species from mating or make it unsuccessful, is the key to speciation (Dobzhansky, 1937). Dobzhansky's contribution included also a description of various barriers preventing gene flow between species (Dobzhansky, 1937)(Dobzhansky, 1951)(Coyne & Orr, 2004) (see Box 5).

Box 5 | Classification of reproductive isolation barriers

I. Premating isolating barriers. Isolating barriers that impede gene flow before transfer of sperm or pollen to members of other species.

A. Behavioral isolation. Includes all differences that lead to a lack of cross-attraction between members of different species, preventing them from initiating courtship or copulation.

B. Ecological isolation. Isolating barriers based primarily on differences in species' ecology, i.e. barriers that are direct byproducts of adaptation to the local environment.

1. Habitat isolation. Species have genetic or biological propensities to occupy different habitats when they occur in same general area, thus preventing or limiting gene sexchange through spatial separation during the breeding season. This isolation can be caused by differential adaptation, differential preference, competition, or combination of these factors.

2. Temporal (allochronic) isolation. Gene flow between sympatric taxa is impeded because they breed at different times.

3. Pollinator isolation. Gene flow between angiosperm species is reduced by their differential interactions with pollinators. This can occur via pollination by different species, or by pollen transfer involving different body parts of a single pollinator species.

C. Mechanical isolation. Inhibition of normal pollination or copulation between two species due to incompatibility of their reproductive structures. This incompatibility may result from lack of mechanical fit between male and female genitalia (structural isolation) or the failure of heterospecific genitalia to provide proper stimulation for mating (tactile isolation).

D. Mating system "isolation". The evolution of complete or partial self-fertilization (autogamy)or the asexual production of offspring (apomixis) that can result in the creation of a new taxon or set of lineages. This is not an isolating barrier in the same sense as the others in this list.

II. Postmating, prezygotic isolation barriers. Isolating barriers that act after sperm or pollen transfer but before fertilization.

A. Copulatory behavioral isolation. Behavior of an individual during copulation is insufficient to allow normal fertilization.

B. Gametic isolation. Transferred gametes cannot effect fertilization

1. Noncompetitive gametic isolation. Intrinsic problems with transfer, storage, or fertilization of heterospecific gametes in a single fertilization between members of different species.

2. Competitive gametic isolation. (conspecific sperm or pollen preference) Heterospecific gametes are not properly transferred, stored or used in fertilization only when competing with conspecific gametes.

Box 5. Classification of reproductive isolation barriers (continues)

III. Postzygotic isolating barriers (hybrid sterility and inviability)

A. Extrinsic. Postzygotic isolation depends on the environment, either biotic or abiotic.

1. Ecological inviability. Hybrids develop normally but suffer lower viability because they cannot find an appropriate ecological niche.

2. Behavioral sterility. Hybrids have normal gametogenesis but are less fertile than parental species because they cannot obtain mates. Most often, hybrids have intermediate phenotypes or courtship behaviors that make then unattractive.

B. Intrinsic. Postzygotic isolation reflects a developmental problem in hybrids that is relatively independent of the environment.

1. Hybrid inviability. Hybrids suffer developmental difficulties causing full or partial lethality.

2. Hybrid sterility.

a. Physiological sterility. Hybrids suffer problems in the development of the reproductive system or gametes.

b. Behavioral sterility. Hybrids suffer neurological or physiological lesions that render them incapable of successful courtship.

Based on (Dobzhansky 1951), modified by (Coyne & Orr 2004)

Allopatric speciation

One of the most pronounced barriers to gene flow is geographical isolation. Under an allopatric model of speciation, the levels of divergence between geographically isolated populations, increase over time. Depending on how much time has passed since the separation, the subsequent reunion of members of the populations will follow one of the three scenarios: 1) with high level of similarity the individuals will merge back into one single population/species; 2) with some level of divergence hybrids may be formed, but will be affected by lower survival rate or fertility; 3) with high level of divergence the hybridization is arrested or leads to sterile hybrids (Maynard Smith, 1978a). The barriers of fertilization, which arose without the contact between new forming species, cannot be due to selection, but rather random genetic drift. Dobzhansky's studies on male hybrid sterility in *Drosophila* provided initial insights on the number of interacting genes and contribution of chromosomes, especially sex chromosomes, to gene flow arrest (Dobzhansky, 1936).

In the second scenario, if hybridization results in low fitness the offspring, natural selection will act to strengthen reproductive barriers and prevent costly mating (Dobzhansky, 1951); this evolutionary process is called "reinforcement". Recent theoretical models identified optimal conditions for reinforcement to occur (Kelly & Noor, 1996)(Kirkpatrick & Servedio, 1999)(Servedio, 2000) and some empirical indications were found in *Drosophila* species (Coyne & Orr, 1989)(Noor, 1995)(Ortiz-Barrientos *et al.*, 2004).

Speciation and mate recognition proteins

Nowadays, genomic techniques are a source of data on barriers to the gene flow that underlies evolution of new species. By studying laboratory crosses and natural populations comparative genomics analysis may provide general patterns and rules that contribute to speciation and answer the questions of relative frequency of these processes in nature.

Although the genetic bases of reinforcement are still under investigation, it seems logical that genes related to mate recognition would be the preferential choice for selection to act on. Ortiz-Barrientos and colleagues provided the first genetic study of adaptive female preference in *Drosophila pseudoobscura,* showing multiple genes potentially involved in mating discrimination, including olfactory receptors (Ortiz-Barrientos *et al.*, 2004). Additionally, many species display rapid evolution of gamete recognition genes which may contribute to speciation (Willie J Swanson & Vacquier D Victor, 2002)(Swanson *et al.*, 2001b)(Swanson & Vacquier, 2002)(Panhuis *et al.*, 2006)(Clark *et al.*, 2006). The reason behind the rapid evolution of only a subset of genes compared to the rest of the genome, however, raises difficult questions. Given that successful mating ensures species continuity, why would reproductive genes evolve so rapidly risking changes that may lead to gamete sterility.

Rapid evolution of gamete recognition genes

It is possible that a group of individuals will display preferences for mating as a consequence of the appearance of new heritable sexual traits (Noor & Feder, 2006). These novel interactions may arise on the ground of morphology and behavior (new ornamentation in a male individual that is recognized by a choosy female) or on the molecular level of gamete recognition (a variant of sperm ligands recognized by a variant egg receptor). Sex genes on the fast evolution route could also create opportunities for the population to occupy novel niches. If new variants of recognition genes arose together with a phenotype adapted to a particular niche, this would allow for expansion of such trait and its fixation (Coyne & Orr, 2004)(Goodenough *et al.*, 2007). Consequently, one would expect more speciose lineages to have rapidly evolving sex-related genes (Fig.2). However, more research is needed to test this hypothesis in nature.

An alternative explanation for the rapid evolution of reproductive proteins is provided by sperm competition, sexual selection and sexual conflict. Sperm competes with the other sperm to be the first one to fertilize the egg, so selection may favor fast swimmers or best binding properties to the egg surface (Clark *et al.*, 1999). However, cryptic female choice may favor a specific allele of sperm-surface protein (Eberhard, 1996)(Palumbi, 1999). This may result in sexual conflict, especially if sperm is present in large quantities (Rice & Holland, 1997). The threat of polyspermy may select for mechanisms preventing fast or multiple fusions with sperm in egg cells, including adaptations in the egg coat proteins (Swanson *et al.*, 2001b).

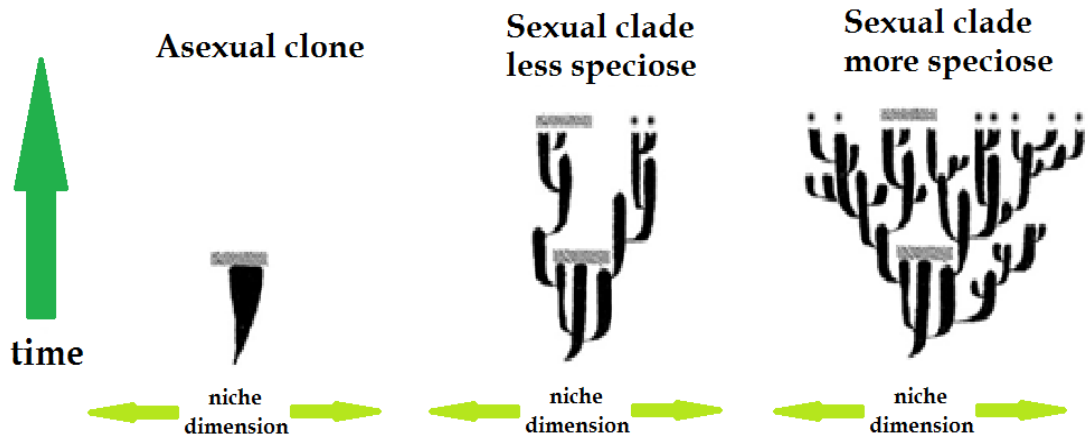


Figure 2. Graphical representation of evolutionary patterns and the influence of speciosity on niche dimension (after Stanley (J. A. Coyne & H. A. Orr 2004)). Gray bars denote events that wipe out major niches; circles indicate extant members of a clade.

Marine broadcast spawning species serve as great source of information about fertilization mechanisms, and structure and evolution of gamete recognition genes (Swanson & Vacquier, 2002). Because fertilization takes place in the water column, pre-zygotic incompatibility is the physical barrier that prevents crosses between sympatric species (Levitan *et al.*, 2004). Thus rapid evolution of the reproductive proteins is under positive (directional) selection, increasing the proportion of nonsynonymous/synonymous substitutions (Swanson & Vacquier, 2002). A Sexually Induced Gene 1 (Sig1) in a diatom *Thalassiosira* spp. was shown to have high divergence, both within and between species (Armbrust & Galindo, 2001). The gene is upregulated during mating (Armbrust, 1999), however its exact function is not known. Nevertheless, its extreme divergence indicates a possible function as a barrier to hybridization between geographically distant strains (Sorhannus, 2003)(Sorhannus & Kosakovsky Pond, 2006). Another example comes from abalone (*Haliotis*) and turbin snails (*Tegula*). During mating, these gastropods release a soluble protein, lysin, from the sperm acrosome to create a hole in the egg envelope through which the sperm can reach the egg cell membrane. The lysin amino-acid sequences from different species are extremely divergent and there is evidence that this divergence is a result of adaptive evolution (Hellberg & Vacquier, 1999)(Metz *et al.*, 1998)(Yang *et al.*, 2000b). Moreover, a protein that mediates egg-sperm fusion in abalone (protein sp18) (Swanson & Vacquier, 1995b) showed up to 73% divergence in amino acids between five Californian species (Swanson & Vacquier, 1995a). It is remarkable that exons of these genes seem to evolve 20 times faster than introns (Metz *et al.*, 1998). Abalone lysin and sp18 are probably the fastest evolving proteins known so far and

the most robust example of positive Darwinian selection (Swanson & Vacquier, 2002). It was suggested that abalone lysin and its multi-repeat domain containing egg receptor (VERL) diversify through concerted evolution to maintain recognition between lysin and its constantly changing female receptor (Fig. 3) (Swanson & Vacquier, 1998)(Swanson *et al.*, 2001a).

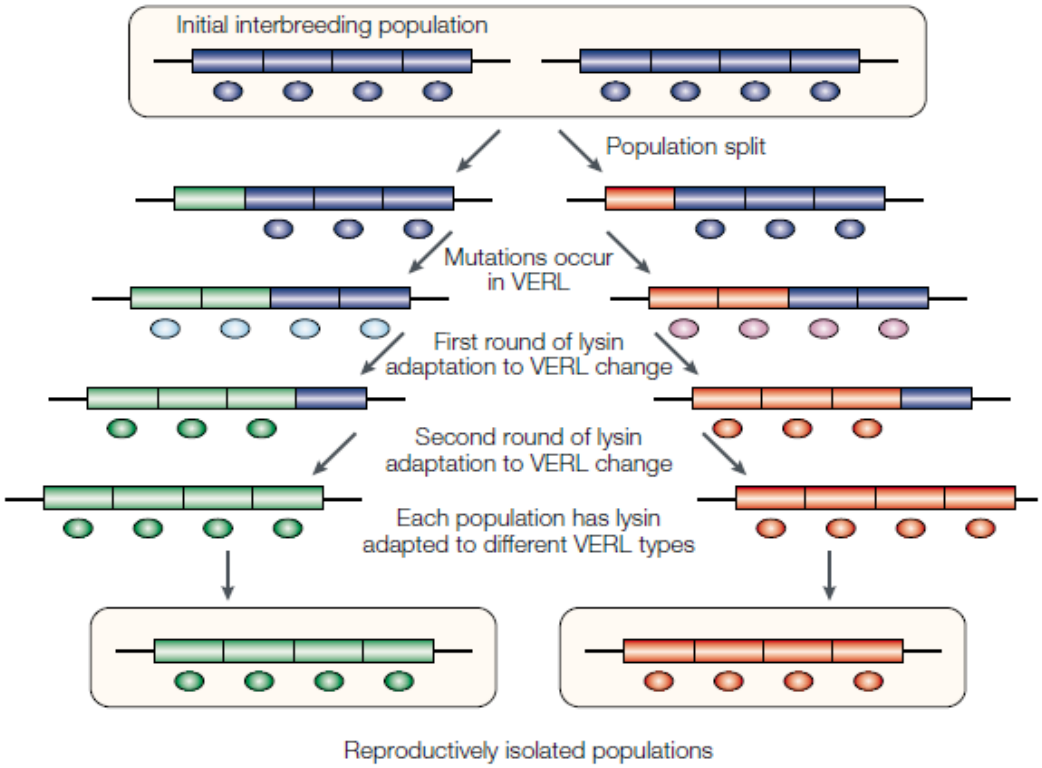


Figure 3. Hypothetical scenario of coevolution of lysine and VERL in abalone (after Swanson and Vacquier in (W. J. Swanson & V. D. Vacquier 2002)). Lysin is represented as circles and VERL is shown as a bar, divided into repeat domains. A mutation that changes one domain repeat in VERL might not interfere with lysin binding to remaining repeats units. However, if the domain becomes more prevalent through i.e. unequal crossing over and gene conversion, lysin will adapt to this change to maintain an efficient VERL–lysin interaction. It might take multiple rounds of adaptation for lysin corresponding to one change in VERL.

One possible scenario of how reproductive proteins may evolve is shown in Figure 3. The vitelline envelope receptor for lysin (VERL) is a fibrous molecule of 1,000 kDa containing approximately 28 repeats of 153 amino acids (Swanson & Vacquier, 1997)(Swanson & Vacquier, 1998). VERL repeats are homogenized in the sequence by unequal crossing over and gene conversion (Elder & Turner, 1995) resulting in repeats that are more similar within

a species than between species. VERL repeats did not show sign of positive selection, however the redundant nature of the VERL might alleviate the functional constrains of each repeat and result in relaxed selection on repeats (Swanson & Vacquier, 1998). It appears that lysin is evolving to match changes in a neutrally drifting VERL (Swanson *et al.*, 2001a).

It was demonstrated in many species, including *Chlamydomonas* (Goodenough *et al.*, 2007), that rapidly evolving sex-related genes carry regions of low amino-acid complexity (LCRs) that are vulnerable to insertions/deletions. Other sex-related genes possess substitution-resilient, secondary structures that persist despite the accumulation of numerous amino-acid substitutions (Goodenough *et al.*, 2007). These structures would facilitate accumulation of genetic changes. However, functional studies are necessary to resolve the consequences of the rapid evolution of reproductive proteins.

4 Algae as a model group for sexual reproduction studies

Algae show an impressive range of variability regarding their life cycles and reproduction strategies. Haploid and diploid generations may exist independently connected by meiosis and syngamy, with some algae persisting only in haploid (e.g. most green algae) and some only in diploid (i.e. diatoms, *Fucus*) states. Sexuality is defined by dioecism, monoecism, isogamy, anisogamy and oogamy depending on the species.

Early embryogenesis

Algal systems have been used to address a large number of fundamental questions in developmental biology (Quatrano, 1990). Their mechanisms for differentiation and development have been studied for more than 100 years, providing advances in the three main processes related to early embryogenesis: gamete attraction and recognition during fertilization (Bolwell *et al.*, 1979)(Boland *et al.*, 1983)(Müller & Schmid, 1988)(Schmid, 1993), hormone-mediated control of gamete formation, fusion and zygote development (Lewin, 1956)(Basu *et al.*, 2002)(Le Bail *et al.*), and establishment of cell polarity (Kropf, 1992)(Brownlee *et al.*, 2001). Recent algal genetic studies present valuable insight into the evolution of mating loci (Ferris & Goodenough, 1997)(Ferris *et al.*, 2010), transition towards anisogamy (Umen, 2011)(Togashi *et al.*, 2012) as well as genetic control of the cell cycle and gametogenesis (Gillard, 2009)(Coelho *et al.*, 2011)(Miller *et al.*, 2010). Since reproductive isolation in free-spawning algae may rely on small changes in the gamete recognition proteins that control fertilization, it is also possible to study gamete incompatibility in relation to speciation and sexual selection (Ferris *et al.*, 1997)(Armbrust & Galindo, 2001).

Challenges of external fertilization

The marine environment poses challenges for fertilization, because gametes are released directly in the water where they need to attract, recognize and fuse with the right partner. Thus, algae developed various mechanisms to control gametogenesis and synchronize gamete release depending on environmental factors. These factors may include water temperature, tide level, day length and lunar phases (Lobban & Wynne, 1981) as well as bioactive compounds such as pheromones (see (Boland, 1995) for a review).

Additionally, fertilization in marine organisms is controlled by chemical communication operating at the surface membranes of sperm and eggs. Gamete recognition factors embedded in cell membranes have been at least partially identified for a variety of algal species (Schmid, 1993)(Kim & Kim, 1999)(Kalshoven *et al.*, 1990). A cDNA subtraction library identified a novel Sexually Induced Gene (Sig) family in the centric diatom *Thalassiosira weissflogii* (Armbrust, 1999). The Sig1 gene of this family has been hypothesized to encode a protein involved in gamete recognition. Additionally, the high levels of sequence divergence (both within and between species) place it among rapidly evolving sex related genes found in marine gastropods (Armbrust & Galindo, 2001).

Evolution of mating types

Chlamydomonas reinhardtii is a model organism for studies concerning mating type determination and evolution (Goodenough, 1995)(Ferris & Goodenough, 1997)(Goodenough *et al.*, 2007)(Beck & Haring, 1996)(Umen, 2011). Mating types in this green isogamous alga are governed by genes in the mating locus (MT) and additional autosomal genes (Ferris & Goodenough, 1997), which are induced by nitrogen starvation to enter the sexual reproduction stage (Beck & Acker, 1992). A close multicellular relative of *Chlamydomonas*, *Volvox carteri*, evolved oogamy and has undergone remarkable expansion of its MT locus. Transcriptome analysis of the *Volvox* MT genes revealed more complex gametic expression with multiple sex-dependent transcripts and gender specific selection (Ferris *et al.*, 2010)(Umen, 2011), opening possibilities to investigate co-evolution of sex determining loci with acquisition of multicellularity and major changes in sexual reproduction strategies.

New algal models are emerging in the heterokont phylum, which is phylogenetically distant to both green and red algae (Baldauf, 2003). Much research has been devoted to *Ectocarpus siliculosus,* a model organism for brown algal lineage (Charrier *et al.*, 2008), including studies on gamete recognition (Schmid, 1993), pheromone production (Muller *et al.*, 1971), early development and transition of generations (Arun *et al.*, 2013)(Coelho *et al.*, 2011)(Le Bail *et al.*, 2008). The release of the *Ectocarpus* genome sequence in 2010 (Cock *et al.*, 2010), opened a new window of opportunities in the algal sexual reproduction studies.

Sex mating loci are also being studied in diatom species. These microalgae with a diplontic life cycle show interesting variation in reproductive features ranging from oogamy in homothallic centric diatoms to isogamy in mostly heterothallic pinnate diatoms (Edlund & Stoermer, 1997)(Chepurnov *et al.*, 2004). Moreover, the traditional evolutionary pathway from isogamy through anisogamy to oogamy seems to appear in reverse order in this group, linking morphologically alike gametes in most pennates to ancestral oogamy in centrics (Edlund & Stoermer, 1997)(Chepurnov *et al.*, 2008). With the recent discovery of the genetic basis for mating type determination in the pinnate diatom *Seminavis robusta* (Vanstechelman *et al.*, 2013)(Vanstechelman, 2013), diatom research offers promising advancements in the understanding of evolution and mechanisms of mating system transitions.

5 Aims and outline of the thesis

In this thesis we address the complex matters of sexual reproduction and speciation in brown algae. Using *Ectocarpus siliculosus*, an emerging brown algal model organism, we perform a comparative analysis of isogametes at the commencement of sexual dimorphism evolution. Recent advancement in transcriptome sequencing technologies makes such goals more achievable and the outcome more comprehensive than was previously possible. Gamete transcriptomes are therefore investigated to uncover molecular bases of sexual differentiation. We search for the molecules involved in gamete recognition during fertilization using gene expression data and biochemical experiments. Ultimately, we put to the test the paradigm of rapid evolution of sex-biased proteins, exploring divergence of male putative signaling receptors.

Chapter 2 describes the biology of a brown algal model species *Ectocarpus siliculosus* focusing on sexual reproduction (**Part 2.1**) and gamete surface antigens (**Part 2.2**). In the second part we discuss the current knowledge of gamete recognition proteins and relate to our efforts in isolating gamete agglutinins from *Ectocarpus*. The aim of this chapter is to identify male and female proteins mediating cell-cell recognition in this alga based on partial characteristics provided by Schmid et al. (1993).

In **Chapter 3**, we describe the procedure of analyzing Next Generation Sequencing data. The characteristics of sequencing techniques, data quality and different stringency modes in sequence alignments are presented. This chapter provides an overview of a data analysis workflow leading to a comparative study of *Ectocarpus* gamete transcriptomes (Chapter 4).

Chapter 4 provides the first insight into transcriptomes of protist gametes, revealing many gender-specific and sex-regulated transcripts connected to pheromone production, cell movement, cell wall biosynthesis and others. We also discuss interesting candidates for signaling proteins that might be involved in gametes recognition. In this chapter we aim to unveil hidden genetic differentiation in morphologically identical gametes and hypothesize on the function of sex-biased genes.

In **Chapter 5**, a study of signature of positive selection in potential male gamete receptors is presented. Here we address the divergence of sex-biased genes and discuss its implication in reproductive isolation and speciation.

Finally, general conclusions about this research project and perspectives for the future studies are described in **Chapter 6**.

Chapter 2

Ectocarpus siliculosus as a model organism for sexual reproduction studies

A.L. contributed to this chapter by maintaining algal cultures, performing fertilization assays, protein extractions and chromatography experiments.

Part 2.1 Reproductive biology of *Ectocarpus siliculosus* – overview

1 Introduction to brown algae

Brown algae (Phaeophyceae) are a group of mainly marine, photosynthetic organisms, belonging to the Heterokontophyta (Baldauf, 2003). With a long history of ecological and commercial importance, they are a valuable capital for various bio-industries including food, fertilizers, cosmetics and pharmaceuticals (McHugh, 2003). Ecologically, many large brown seaweeds form forest-like shelters for many marine species and are key primary producers in the coastal regions (Graham *et al.*, 2007).

1.1 Brown algal evolution

Phototrophic organisms show extraordinary divergence in form and function, making it hard to resolve their evolutionary history. There is ample evidence, that primary plastids were derived through an endosymbiotic event between a free living eukaryotic host and a cyanobacteria (Archibald, 2009), giving rise to red, green and glaucophyte algae (Palmer, 2003)(Reyes-Prieto et al., 2007). Understanding plastid acquisition in groups with red-algal derived plastids largely remains an open question. A recent hypothesis holds that a secondary endosymbiotic event with an unidentified and probably extinct red alga, resulted in the plastid of a common ancestor of cryptophytes and haptophytes (Rice & Palmer, 2006)(Sanchez-Puerta & Delwiche, 2008)(Archibald, 2009). However, the spread of plastids from this group to the stramenopile (hosting brown algae) and alveolate lineages by tertiary endosymbiosis, and whether this acquisition was independent or from a common ancestor, remains an open question (Archibald, 2009)(Elias & Archibald, 2009)(Archibald, 2012). Additionally, nuclear genomes of diatoms *Thalassiosira pseudonana* and *Phaeodactylum tricornutum* (Stramenopiles) revealed the presence of both red and green algal genes (Bowler *et al.*, 2008), suggesting a cryptic secondary endosymbiosis event involving green algae before red algal plastid was assimilated (Moustafa *et al.*, 2009). More genomic sequences of both, plastid donors and chromalveolates, are needed to confidently resolve the history and origin of photosynthetic protist lineages.

Unique among the heterokonts, brown algae have acquired complex multicellularity. As a result of this long independent evolution, a genomic repertoire distinct from other

multicellular organisms (e.g. land plants, animals and fungi) has evolved in brown algae. Meanwhile, the core biological processes share enough similarities with plant and animal systems, making it possible to use brown algae as model organisms to address the underlying processes controlling development, embryogenesis, cell polarity and fertilization.

1.2 Life histories and fertilization modes

As a group, Phaeophyceae exhibit complex evolutionary patterns in life histories and fertilization modes. Life cycles of brown seaweeds range from haploid-diploid cycles with sporophytic and gametophytic generations, through diploid life cycles with only one multicellular stage (Bell, 1997) (Fig. 1a). Moreover, sporophytes and gametophytes in haploid-diploid life cycles can be characterized by morphological and size differentiation (heteromorphic life cycle). Mating systems range from isogamy (equal sized gametes) to oogamy (large, non-motile egg and small motile sperm), with many taxa representing an intermediate state of anisogamy (both gamete types are motile, but with clear size differentiation) (Fig. 1b). As it is shown by phylogenetic analysis, distant reproductive characteristics have been changing course multiple times during the evolution of brown algae (Silberfeld *et al.*, 2010)(Fig. 1). However, more sampling, including early diverging lineages, is needed to assess whether these transitions included backward conversions from heteromorphic to isomorphic life histories (Fig. 1a) and from oogamous to isogamous fertilization modes (Fig. 1b) (Silberfeld *et al.*, 2010). This amazing plasticity of reproductive characters makes Phaeophyceae a desired subject of life cycle and reproduction evolution studies.

The release of eggs and sperm in bulk quantities into surrounding sea water accounts for an easy and accessible system to address fertilization. Moreover, examinations of closely related species, which display different reproductive strategies, enable the investigation of evolution of sexes and sexual dimorphism. Nevertheless, the fundamental differences between sexes and mating types are not fully resolved (Charlesworth, 2013). The most common criterion discriminates between large female egg and small motile sperm, using anisogamy as an indication of having two different sexes. However, haploid gametophytes in several dioecious brown algae produce gametes of diverse physiology and behavior, without size dimorphism. These species might therefore be classified as having a form of anisogamy different from other dioecious species (Charlesworth, 2013) and are an interesting source to study evolution of functional diversification in the absence of sexual dimorphism (Fig.1).

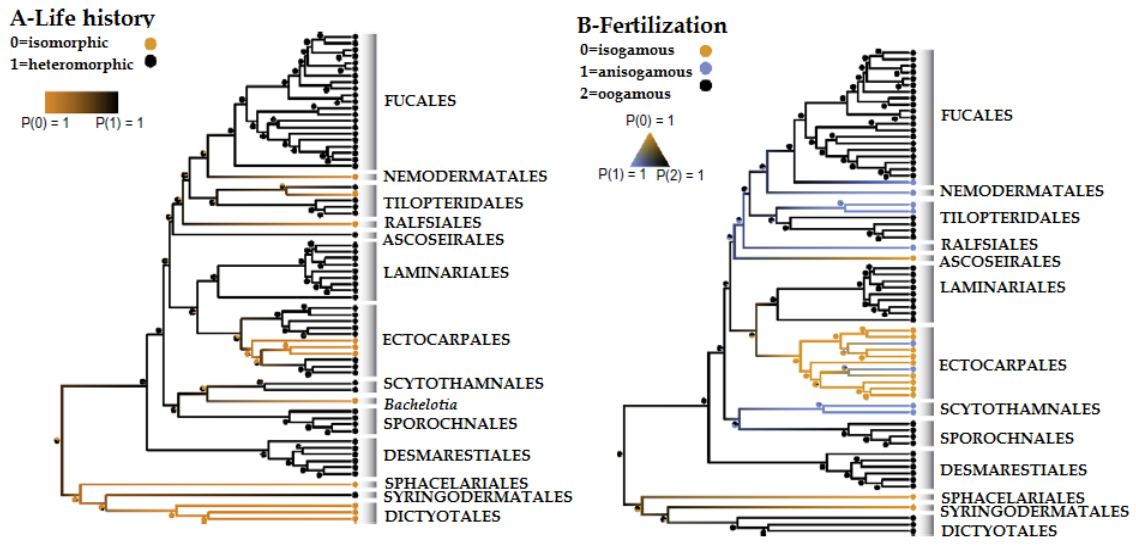


Figure 1. Phylogenetic relationships between phaeophyceaen orders (after Silberfeld et al. (Silberfeld *et al.*, 2010)) with marked life histories (A) and fertilization modes (B).

Successful fertilization in species with freely released gametes depends primarily on the simultaneous release of gametes and an effective attraction/recognition system, favoring mating of gametes of the same species as opposed to interspecific mating. This system is presumed to be under selective pressure, considering that the plethora of co-existing species in marine environment increase the risk of hybrid formation. Gamete recognition genes and pheromone structures demonstrate high divergence and discrimination between marine species (Springer *et al.*, 2008)(Swanson, 1998)(Lessios *et al.*, 2012)(Mah *et al.*, 2005)(Boland, 1995), linking prezygotic gamete incompatibility to speciation (Palumbi, 1994)(Palumbi, 2008). However, the influence of particular genetic changes on the evolution of reproductive isolation and speciation are still poorly understood. To investigate the relationships between isogamy, gamete functional diversification, evolution of sex-related genes and speciation we turned to a brown algal model species *Ectocarpus siliculosus*.

2 *Ectocarpus siliculosus as a model organism for sexual reproduction studies*

Ectocarpus possesses multiple features that make it a desirable model organism in reproduction studies. This filamentous, isogamous alga can complete its life cycle within three months in laboratory conditions, maintained as axenic, unialgal culture. Furthermore, *Ectocarpus* becomes easily fertile in culture and produces a large number of zoids. Genetic crosses are also rather easy to perform. Most importantly, many detailed characteristics have been described for *Ectocarpus* spp., including its life cycle (Müller, 1967), sexual pheromone signaling (Müller *et al.*, 1971)(Müller & Schmid, 1988), cell-cell recognition (Schmid, 1993),

phylogeography (Stache Crain *et al.*, 1997) and sexual compatibility of geographically distant strains (Stache, 1990)(Muller, 1988)(Muller, 1976)(Muller, 1979a)(Muller & Eichenberger, 1995)(Peters *et al.*, 2010b). In 2010 a complete genome sequence of *Ectocarpus* has been released (Cock *et al.*, 2010)(Sterck *et al.*, 2012) and many genetic and genomics tools have followed including a linkage map (Heesch *et al.*, 2010) and microarray design (Dittami *et al.*, 2009) with tools to study gene expression (Dittami), allowing to address a wide range of questions at the molecular level. With the data available, we are now able to study diversification of gamete receptors in relation to the current species concept and known compatibility barriers, incipient gamete differentiation on the whole transcriptome level, and evolutionary variation of sex biased genes in populations and species of *Ectocarpus*.

2.1 Life cycle of Ectocarpus

Multicellular eukaryotes evolved a balanced transition between haploid and diploid growth in their life cycles. Understanding these two generations is an important step to the comprehension of their maintenance and a clue to understand the life cycle itself. The *Ectocarpus* life cycle involves the alternation of slightly heteromorphic haploid and diploid generations, with meiosis preceding the haploid state and syngamy reconstituting the diploid genome (Müller, 1967)(Müller, 1977). Reproduction is both, sexual and asexual (Müller, 1967). Sexual reproduction appears in the form of isogamy, where flagellated gametes are morphologically, but not physiologically, identical (Müller, 1967). Gametes are produced through several mitotic divisions, by haploid male and female gametophytes, in plurilocular reproductive organs. Female gametes settle quickly and start the release of a pheromone (ectocarpene) attracting male swimmers (Müller *et al.*, 1971)(Müller & Schmid, 1988). Gamete recognition in *Ectocarpus* is presumably mediated by lectin-glycoprotein complexes associated with gamete outer membranes (Schmid, 1993). Gamete fusion gives rise to a diploid sporophyte, which can reproduce asexually by diploid zoids produced in plurilocular sporangia or make a transition to the haploid gametophyte phase (Müller, 1967)(Baker & Evans, 1973a). Meiospores result from single meiosis followed by several mitotic divisions in a unilocular sporangium (half of the meiospores are male and half are female) (Baker & Evans, 1973b)(Müller, 1991). Sex determination in *Ectocarpus* is controlled by an ancient UV pair of sex chromosomes (Ahmed et al., unpublished data). Interestingly, unfused gametes develop into parthenosporophytes, which are capable of endoreduplication and reproduce themselves asexually through zoids or produce meiospores and regenerate the gametophyte generation of the same gender (Bothwell *et al.*, 2010). The basic life history of *Ectocarpus* is illustrated in Figure 2.

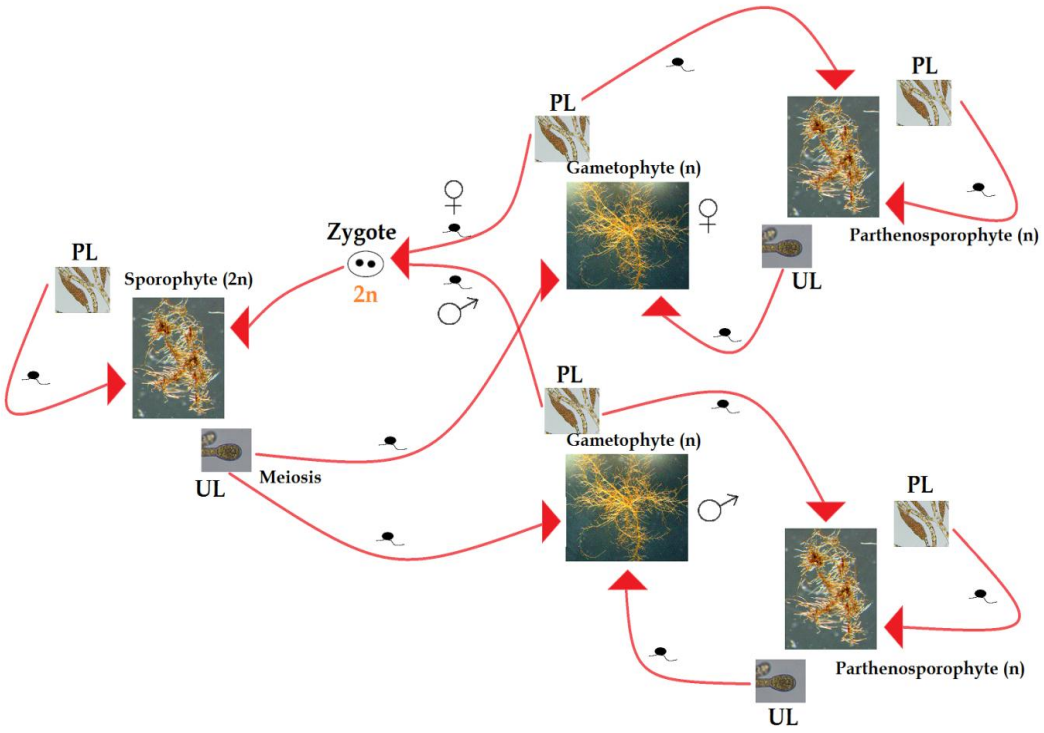


Figure 2. The life cycle of *Ectocarpus siliculosus*. Male and female gametophytes are produced by a diploid sporophyte in unilocular sporangia (UL) during meiosis. Gametophytes produce gametes in plurilocular gametangia (PL). The sexual cycle is completed by fusion of haploid male and female gametes and formation of a diploid sporophyte. Unfused gametes may grow parthenogenetically and later restore the gametophyte generation. Both sporophytes and parthenosporophytes can reproduce asexually by mitospores produced in plurilocular sporangia.

Transition of generations

The sporophytic and gametophytic life stages in *Ectocarpus* can in theory be recognized by their dissimilar morphologies; however architectural plasticity is a well-known phenomenon in this alga (Ravanko, 1970)(Le Bail *et al.*, 2008) which may cause ambiguities when examining field material. Sporophytes (as well as parthenosporophytes) develop from as prostate base and produce less lateral branches, whereas gametophytes are more richly branched and deprived of a prostrate base (Müller, 1964)(Müller, 1967)(Le Bail *et al.*, 2008). In laboratory culture, the two stages can be readily distinguished by two different patterns of early development involving symmetrical (sporophyte) or asymmetrical (gametophyte) first cell division. Analysis of a developmental sporophyte mutant (*immediate upright imm* mutant) which exhibited gametophyte-like asymmetrical division and upright filament indicated that this morphology is controlled by a single-locus Mendelian allele (Peters *et al.*, 2008). Despite the visual resemblance of the gametophyte, the *imm* mutant remained a

functional sporophyte which led to further studies on genes governing the transition of generations.

The discovery of another mutant *ouroboros* (*oro*) resolved the genetic basis for the gametophyte-to-sporophyte life cycle transition. The *oro* mutation was described as a single, recessive, Mendelian locus, providing a unique opportunity to study the effect of a single mutation acting on the level of a whole organism (Coelho *et al.*, 2011). The *oro* mutant is a functional gametophyte producing gametes that continuously renew the gametophyte generation. Similar to the *imm* mutant, *oro* displayed a gametophytic morphology with expression of genes characteristic to gametophytes; however the two loci were shown to be unlinked with *ORO* being at least partially epistatic to *IMM* (Coelho *et al.*, 2011).

Interestingly, a non-cell autonomous transition from the gametophyte to the sporophyte generation can be induced in meiospores carrying a functional copy of the *ORO* gene by a diffusible factor produced by the sporophyte (Arun *et al.*, 2013). These alternated gametophytes develop into functional sporophytes, indicating that the sporophyte factor and *ORO* are part of the same pathway, with ORO acting downstream of the diffusible factor.

These studies enhanced our understanding of the functioning of the gametophyte and sporophyte on the molecular level by generating comprehensive transcriptomic data, identifying many differentially expressed regulatory genes (Coelho *et al.*, 2011). Furthermore, they decoupled ploidy from life cycle stage, producing diploid gametophytes in double *oro* mutants and indicated that the male haplotype of the sex determining locus is dominant over females (Coelho *et al.*, 2011). Although explaining the maintenance of haploid-diploid life cycles is problematic, it is now possible to study the relationship between different morphology and adaptation to ecological niches in *Ectocarpus*.

2.2 Gamete's characteristics

Brown algal gametes can exhibit high morphological differentiation, probably in response to the specific requirements of the reproduction (Kawai, 1992). Because male and female swarmers in *Ectocarpus* are morphologically identical and resemble the zoospores (Kawai, 1992), they have been classified as the 'primitive type'. However, subtle differences can be noticed upon close examination of unilocular and plurilocular (asexual and sexual) zoids. Plurilocular zoids are smaller and swim faster with rapid changes of direction comparing to unilocular zoids, which are also distinguished by physical separation of the nucleus from the chloroplast (Baker & Evans, 1973a)(Baker & Evans, 1973b). Moreover, unilocular zoids are characterized by lower secretory activity (Baker & Evans, 1973b). Furthermore, a British *Ectocarpus* strain demonstrated substantial size difference discriminating gametes (4-5μm in diameter) from asexual spores (7-10 μm in diameter)(Müller, 1977).

Ultrastructure

Ectocarpus gametes are characterized by a single chloroplast with a pyrenoid, an eyespot and two flagella; a short, naked, posterior flagellum and a longer anterior flagellum with mastigoneme hairs (Baker & Evans, 1973a)(Lofthouse & Capon, 1975)(Maier, 1997a). The general structure of the cell is shown in Fig. 3. Zoids are capable of positive phototaxis and their posterior flagella display strong autofluorescence under blue light (Kawai *et al.*, 1990). The large eyespot, visible under light microscopy, encloses the dorsal swelling of a posterior flagellum, and contains densely packed osmiophilic droplets (Maier, 1997a)(Maier, 1997b)(Fu *et al.*, 2013). The flagellar swelling accommodates finely granular material, concentrated towards the eyespot near the flagellar membrane and on the dorsal side of the swelling (Maier, 1997a). The presumable function of the eyespot is to reflect and focus light on to the photoreceptor (Kawai *et al.*, 1990)(Kreimer *et al.*, 1991), which seems to involve flavin as the photoreceptor pigment (Müller *et al.*, 1987).

Electron microscopic studies of male gametes indicated a high proportion of heterochromatin in the nucleus (Maier, 1997a). Additionally the Golgi region showed high activity, with vesicles containing fine fibrillar material in the *cis*-region and electron-dense material in the *trans*-region (Baker & Evans, 1973a)(Maier, 1997a). This secretory activity was connected to biosynthesis of polysaccharides in zoospores; the fibrous material is believed to be an adhesive, binding the cell to the substratum after settlement (Baker & Evans, 1973a). After one or two hours, other vesicles carrying opaque or transparent content extrude the primary cell wall material during germination (Baker & Evans, 1973a). Occasionally, vesicles containing bundles of tubules were found in the *trans*-Golgi sector of male gametes, most likely representing mastigoneme hairs to be deposited on the anterior flagellum (Maier, 1997a).

Flagella

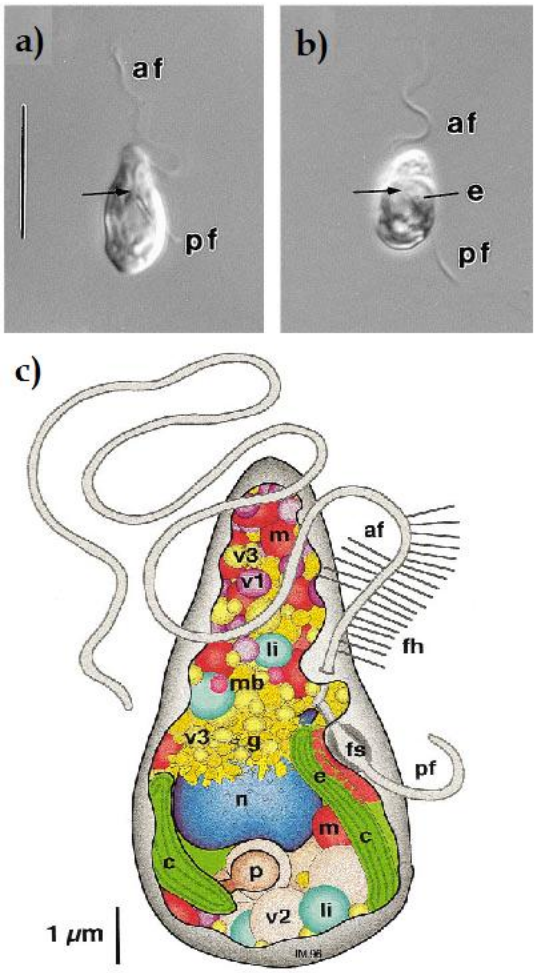
Ectocarpus gametes and spores move around with meandering beats of the anterior flagellum. The posterior flagellum is rather passive, but occasional beats cause changes in swimming direction up to 180° (Geller & Mueller, 1981). Multiple mitochondria are found associated with the anterior flagellar root with outer mitochondrial membranes connected to the microtubules with fine fibrils (Maier, 1997a), probably providing energy for the sliding of microtubules.

In addition to propagating cell movement, the anterior flagellum has an important function in establishing gamete contact during fertilization with its highly sensitive acronema (extended flagellar tip) (Müller & Falk, 1973). Nevertheless, detailed electron microscopy studies of the

flagellar apparatus in both female and male gametes have not found any structural differences (Müller & Falk, 1973)(Maier, 1997b).

2.3 Fertilization

Fertlization success in free spawners depends highly on gamete recognition efficiency. Thus, brown algae have developed a unique, sensitive, chemical communication method as part of the sexual reproduction process. Nonetheless, the chemical structure of the messengers was shown to be limited, meaning that the same molecule may serve as an attractant for gametes of other species or genera (Boland & Mertes, 1985)(Boland *et al.*, 1983)(Boland, 1995).



Abbreviations :

af, anterior flagellum ;

c, chloroplast ;

e, eyespot;

fh, flagellar hairs (present along entire length);

fs, proximal swelling of the posterior flagellum ;

g, Golgi;

li, lipid body;

m, mitochondrion;

mb, microbody;

n, nucleus ;

p, pyrenoid;

pf, posterior flagellum ;

v1, physode; v2, storage vacuole ;

v3, vesicles with presumably cell wall or adhesive material.

Figure 3. Gamete structure of *Ectocarpus siliculosus*.
(a, b) Living gametes of *Ectocarpus siliculosus* (strain PAr-27a). The arrow indicates the point of flagella insertion. Scale bar represents 10µm. (c) Schematic representation of a gamete cell showing the distribution of the cellular organelles (after (Maier, 1997a)).

Pheromone attraction

After settling down, female gametes digest their flagella, adopt a round cell shape and start to produce olefinic hydrocarbons to allure male partners (Muller, 1979b). Ectocarpene, the first algal pheromone described (Müller *et al.*, 1971), composes 90% of a hydrocarbons bouquet in *E. siliculosus*, with other compounds being dictyotene, multifidene, hormosirene and finavarrene (Stratmann *et al.*, 1993). More recent studies indicated that ectocarpene is a product of biodegradation of a thermally labile precursor, pre-ectocarpene, which is a much stronger attractant and the actual pheromone in *Ectocarpus* (Boland, 1995). Male gametes respond to the pheromone by changing their locomotive behavior. They move in clockwise loops, narrowing down the radius to the female cell until fertilization takes place (Geller & Mueller, 1981). However, ectocarpene is also produced by other species of brown algae in the orders Ectocarpales and Sphacelariales (Müller *et al.*, 1985), implying the need for an additional selection mechanism.

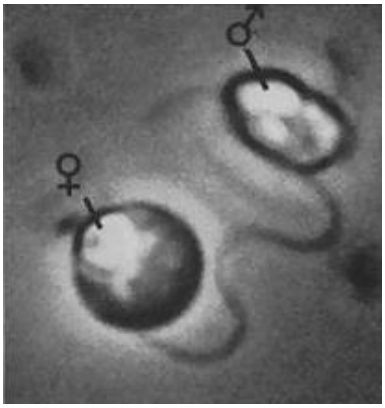


Figure 4. Initial gamete recognition in *Ectocarpus siliculosus.*
Living male gamete is attached by the tip of its anterior flagellum to the surface of a rounded female gamete. (After (Müller & Falk, 1973)).

Gamete surface recognition proteins

Gamete recognition is mediated by N-acetyl glucosamine (GlcNAc) residues exposed on the plasma membrane of the female gametes and a lectin-like receptor on the tip of male anterior flagella (Schmid, 1993)(Schmid *et al.*, 1994b) (Figure 4). This was deduced after fertilization inhibition experiments (Port Aransas strain; PAr 10n female, PAr 27a male) in which wheat germ agglutinin (WGA) inhibited cell-cell recognition when pre-incubated with female gametes (but not male gametes). A similar effect was obtained when male gametes were pre-incubated with GlcNAc, which did not influenced female gametes (Schmid, 1993). Purified outer membranes from female gametes (PAr 10n) also contained gamete specific glycoproteins binding WGA (Schmid *et al.*, 1994b). Nevertheless, efforts of visualizing the recognition moieties on the surface of living cells with fluorescently labeled WGA-lectin did

not differentiate between male and female binding patterns (Maier, 1995) and were thus unsuccessful in localizing sexual receptors *in situ*.

Fertilization success and zygote discrimination

Maximum fertilization success oscillates around 50% and is dependent on the concentration of male gametes (Schmid, 1993). No physiological or genetic aspects are known to be responsible for this low fertilization level; however, it is consistent in other algal species, in which sexual reproduction is not obligatory and unfertilized gametes are able to develop parthenogenetically (Destombe & Cembella, 1990). Zygote formation reaches saturation 20 minutes after gamete mixing and fertilized cells can be recognized by two characteristic eye-spots (Fig. 5a,b)(Schmid, 1993). Moreover, we found that primary cell wall deposition within a few minutes after fusion makes it possible to discriminate zygotes from gametes using Calcofluor staining (Fig. 5c).

Physiological anisogamy

Significant differences between gametes were observed on the protein level, when whole cell extracts were compared (Schmid *et al.*, 1994b). Furthermore, sex correlated fatty acid storage have been found in female gametes, probably as a precursor to pheromone production (Schmid *et al.*, 1994a). Additionally, mitochondria show a strictly maternal inheritance pattern in zygotes; however, plastids are donated by both parents (Peters *et al.*, 2004). These biochemical, behavioral and genetic dissimilarities prove that *Ectocarpus* isogametes, although morphologically identical, are physiologically anisogamous.

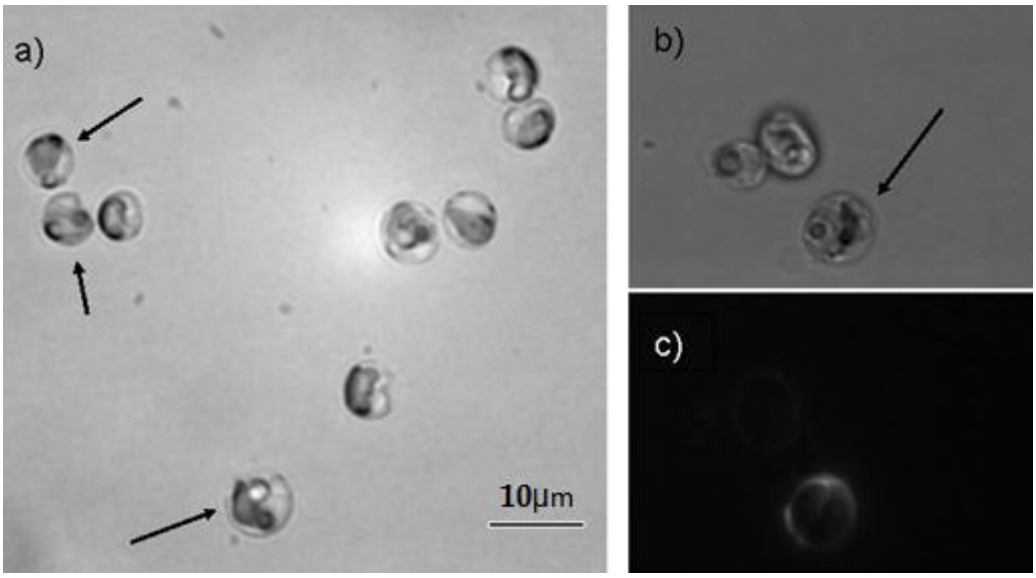


Figure 5. Fertilization in *Ectocarpus siliculosus*. a) Arrows pointing to fertilized eggs; zygotes show 2 eyespots. b) Zygote and 2 unfertilized gametes; c) picture b stained with Calcofluor white, zygote developing primary cell wall.

3 Summary

The *Ectocarpus* genome project together with the numerous genetic, genomic and biochemical tools which are available has provided access to a relatively unexplored branch of the eukaryotic tree of life. Mutant screens, gene expression analysis and possible genetic transformation will allow analyzing gene function and addressing questions in a range of interesting biochemical events that take place during gamete interaction at fertilization, such as chemoreception, cell-cell recognition and fusion processes. All together, the rich history of research (for a review see (Charrier *et al.*, 2008)) and its distinguished life cycle with isogamous fertilization mode, make *Ectocarpus* a valuable model species in reproduction studies in its broad sense.

Part 2.2 Isolation of gamete recognition proteins in *Ectocarpus siliculosus*

1 Introduction

Gamete recognition in sexual eukaryotes is mediated by sex-specific proteins that are unique to the gender and species, which assure that mating occurs between correct individuals (Clark *et al.*, 2009). Sex recognition genes show outstanding divergence, suggesting a unique selection pattern acting upon them during the speciation process (Vacquier, 1998)(Swanson & Vacquier, 2002). Therefore studies of the sex-specific proteins and their evolution may provide a useful insight into species isolation and diversification (Palumbi, 1994)(Palumbi, 2008)(Turner & Hoekstra, 2008).

The fertilization capacity and specificity depends on the functional structure of the recognition molecules, which establish chemical bonds between ligands carrying matching contact sites. Early studies using a unicellular green alga *Chlamydomonas* indicated that such a broad and versatile recognition system can be achieved with lectins and glycoproteins, where one mating partner expresses a glycosylated antigen and the other - a complementary agglutinin (Wiese, 1974)(Weise & Wiese, 1978). This fundamental differentiation grants an efficient gamete asymmetry, comparable to morphological dimorphism in anisogamous species (Weise & Wiese, 1978). Further experiments proved similar interactions were in operation in a plethora of other organisms, from amoebae (Araki *et al.*, 2012) to metazoans (Swanson & Vacquier, 1997)(Loeser & Tulsiani, 1999)(Vilela-Silva *et al.*, 2008) (Fallis *et al.*, 2010)(Clark, 2013).

Fertilization patterns in seaweeds are of great interest from a molecular and ecological perspective. Successful reproduction in many species depends on a delicate equilibrium between biotic and abiotic conditions that trigger synchronous gamete release (including water movement and light) (Brawley *et al.*, 1999)(Pearson & Serrao, 2006) and affect gamete viability, but the first and crucial step of successful reproduction is gamete recognition. Lectin-carbohydrate recognition systems have been reported in marine algae numerous times, however most of the experiments used indirect evidence of inhibition of gamete binding by carbohydrates or plant lectins (Bolwell *et al.*, 1979)(Kim & Kim, 1999)(Schmid, 1993)(Kim & Fritz, 1993). Although several studies have reported on the isolation of marine algal lectins, the number of these proteins that have been purified and characterized is still

small (Mori *et al.*, 2005)(Kim *et al.*, 2006)(Han *et al.*, 2012) and does not include lectins from brown algae.

In a series of papers Christina Schmid (Schmid, 1993)(Schmid *et al.*, 1994b) characterized the affinity of recognition proteins on *E. siliculosus* male and female gametes. The male gametes agglutinin showed affinity to N-acetyl glucosamine (GlcNAc), whereas the complementary female ligands were glycoproteins expressing terminal GlcNAc residues (Schmid, 1993)(Schmid *et al.*, 1994b). This was deduced after a fertilization inhibition experiment in which male gametes pre-incubated with GlcNAc or female gametes pre-incubated with wheat germ agglutinin (WGA), which binds GlcNAc residues, could not fuse and form zygotes. However, the exact genes and proteins involved could not be fully elucidated.

The aim of this study is to isolate and characterize sex related agglutinins in *Ectocarpus* gametes based on fertilization block experiments originally performed by Schmid (1993). We apply chromatography techniques to isolate lectins and glycoproteins based on their binding specificity and mass spectrometry measurements to identify retained proteins. Discovery of gamete recognition proteins would shed light on the complex matter of sexual reproduction and speciation in brown algae.

2 Materials and methods

Algal cultures

Female and male gametophytes of *Ectocarpus siliculosus* (NZKU1-3 male (CCAP 1310/56), NZKU32-22-21 female - obtained from a meiospore of NZKU z32 (CCAP 1318/85); origin Kaikoura, New Zealand) were maintained as unialgal clonal cultures in modified Provasoli enriched natural sea water (West & McBride, 1999). Both gametophytes descend from a single diploid sporophyte collected in Kaikoura, New Zealand representing '*Ectocarpus* lineage 4' according to Stache-Crain *et al*. (1997) (the Ectocarpus sequenced strain . Cultures were kept at 12^0C with 14h light/10h darkness cycles (30 µmol × m^{-2} × s^{-1} flux density). To induce gamete release fertile gametophytes were transferred to Petri dishes with minimal amount of water only and incubated overnight at 4^0C in complete darkness (Coelho *et al.*, 2012). Gamete release was induced by immersing cultures in fresh medium in front of direct light at room temperature. Gametes were collected using a micropipette, transferred into 1.5ml Eppendorf tubes and pelleted at 5,000×g for 5 minutes. Gamete pellets were flash-frozen in liquid nitrogen and stored at -80^0C before protein extraction.

Protein extraction

Gamete membranes were extracted according to protocol of Schmid (Schmid *et al.*, 1992). In short, gametes were homogenized in extraction buffer (25mM Tris/MES pH 7.2, 250 mM

sucrose, 3mM EDTA, 2,5mM DTT, 10% PVP, proteinase inhibitors cocktail (Roche)) and centrifuged at 8000g for 15 minutes followed by ultracentrifugation at 100 000g for 1 hour. Whole gamete extracts (crude extracts)were accomplished by homogenization in hypo-osmotic buffer (50mM Tris-HCl pH 7.2, 1mM EDTA, 1mM DTT, 1% Troton-X, proteinase inhibitors cocktail (Roche)) and incubation at 4°C for 2 hours to overnight. The extract was subjected to centrifugation at 3000g for 10 min and the supernatant was collected for further analysis. All chemicals were obtained from Sigma-Aldrich.

Column chromatography

The affinity columns used in this chapter consisted of Sepharose 4B beads with immobilized WGA or with carbohydrate structures such as N-acetyl D-glucosamine (GlcNAc), mannose, fucose and fetuin. The columns were equilibrated with PBS buffer. Extracts of *E. siliculosus* male and female gametes ($>10^6$ cells per experiment) were passed through the chromatography columns and washed with the washing buffer (PBS) until the absorbance (at 280nm) of the flow-through fraction was lowered to 0.001. Bound proteins were eluted with 20mM 1,3 diaminopropane pH 10 in distilled water and concentrated with Vivaspin columns (GE Healthcare) or precipitated overnight at -20°C with acetone/10%TCA. The protein concentration was determined by the method of Bradford (Bradford, 1976) or measured on a Qubit fluorometer with the Qubit Protein Assay Kit (Invitrogen). Samples were analyzed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE)(Laemmli, 1970). Individual protein polypeptides were excised from the gel, trypsin digested and sequenced using Matrix-Assisted Laser Desorption/Ionization (MALDI) or Liquid chromatography mass spectrometry (Figure 1).

Chemical deglycosylation

Female gamete protein samples eluted from the WGA column were lyophilized overnight and subjected to chemical deglycosylation with TFMS/10%toluene. TFMS (trifluoromethane-sulfonic acid) achieves rapid, non-selective deglycosylation of both N-linked and O-linked glycans, without compromising the protein moiety. Protein samples with 300µl of TFMS / 10% toluene were cooled in an ethanol dry-ice bath and placed at -20°C for 4h. Subsequently, 600µl of a mix (360µl pyridine+120µl MeOH+120µl H_2O) was added to the sample, followed by 0.5% AMBIC. Proteins were precipitated overnight at -20°C with acetone.

Fertilization block assay

The fertilization block assay with lectins was performed according to Schmid (Schmid, 1993). 5µl of freshly released female gametes were placed on a glass slide and allowed to settle in

the darkness for 30 minutes in the moist chamber. Cell density was adjusted to ca. 10^3 cells. Solutions of lectins in sea water were added to the samples in a final concentration of 0.1 mg/ml. The following lectins were used: PNA (Peanut agglutinin), SBA (Soybean agglutinin), WGA (Wheat germ agglutinin), LCA (Lentil agglutinin), GNA (*Galanthus nivalis* agglutinin), DSA (*Datura stramonium* agglutinin) (Table1).

A control assay was performed with mixed gametes in sea water only. Settled female gametes were rinsed with culture medium and 100µl of freshly released male gametes (>10^5 cells μl^{-1}) were added. Gametes were allowed to fertilize for 20 minutes, swimming gametes were rinsed off and Calcofluor white solution was added.

Calcofluor staining

The staining protocol was based on procedure for fixed and fresh tissue by Clark (1981). Calcofluor white stock solution (Sigma-Aldrich) was diluted with natural sea water in 1:100 proportions. Settled gametes were incubated in the Calcofluor dilution for 30 minutes in the darkness and washed three times with sea water before examination under the UV light. As a control of parthenogenic development, samples of settled female gametes were stained to establish the onset of cell wall development without fertilization.

3 Results

3.1 Gamete recognition proteins isolation

Here we report our efforts to purify the sex-specific lectins and glycoproteins from *Ectocarpus siliculosus* (Kaikoura, New Zealand strain). Based on the inhibitory effects of N-acetyl glucosamine and wheat germ agglutinin on fertilization (Schmid, 1993), we chose to use affinity columns consisting of Sepharose 4B with immobilized GlcNAc and WGA. Male and female gametes extracts were loaded on both types of columns to isolate membrane bound proteins. Due to the high sample loss with a membrane purification protocol and the limited amount of gamete material available, we decided to use crude gamete extracts in the subsequent chromatography experiments. The eluted fractions from the columns were analyzed by SDS-PAGE with Coomassie staining followed by mass spectrometry compatible silver staining (Pierce). Protein samples from female gametes retained on a WGA column were also subjected to chemical deglycosylation before mass spectrometry to enhance peptide identification by removing the interfering glycans. Both male and female elution fractions from the GlcNAc columns yielded trace amounts of proteins (1-5µg per column experiment) even if the amount of cells used was increased 10 times (>10^7 cells). Silver staining revealed multiple bands which were excised, digested with trypsin and analyzed on

the MALDI-TOF/TOF spectrometer. No proteins could be identified in the deglycosylated female samples, whereas only one band in the sample obtained from male gametes was assigned to CTP-phosphoethanolamine cytidylyltransferase (Esi0125_0050) (band 4, Fig.2b). Following LC-MS/MS ca. 4,000 peptides were obtained from the sample eluted from the WGA column after loading an extract from female gametes. These peptides could be aligned to five proteins (Esi0043_0035, Esi0340_0006, Esi0026_0113, Esi0101_0018, Esi0212_0047) including a mastigoneme protein and a cell wall building enzyme. No additional proteins could be identified from the male gamete fraction that was retained on the columns.

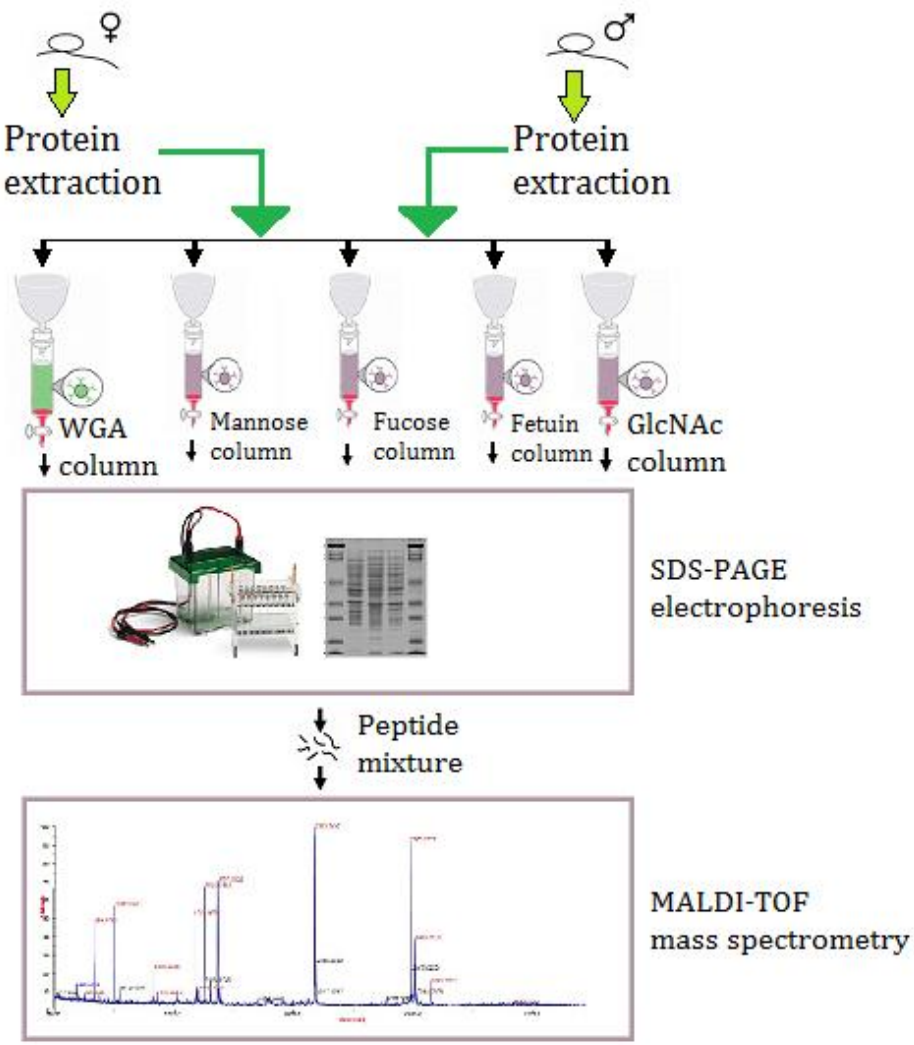


Figure 1. Overview of workflow for the isolation of gamete recognition proteins.

Crude gamete extracts were subjected to column chromatography with immobilized WGA lectin, sugars (mannose, fucose, GlcNAc) or glycoprotein (fetuin). Bound proteins were eluted and separated by size using SDS-PAGE. Protein bands were digested with trypsin and peptides were identified using MALDI-TOF mass spectrometry.

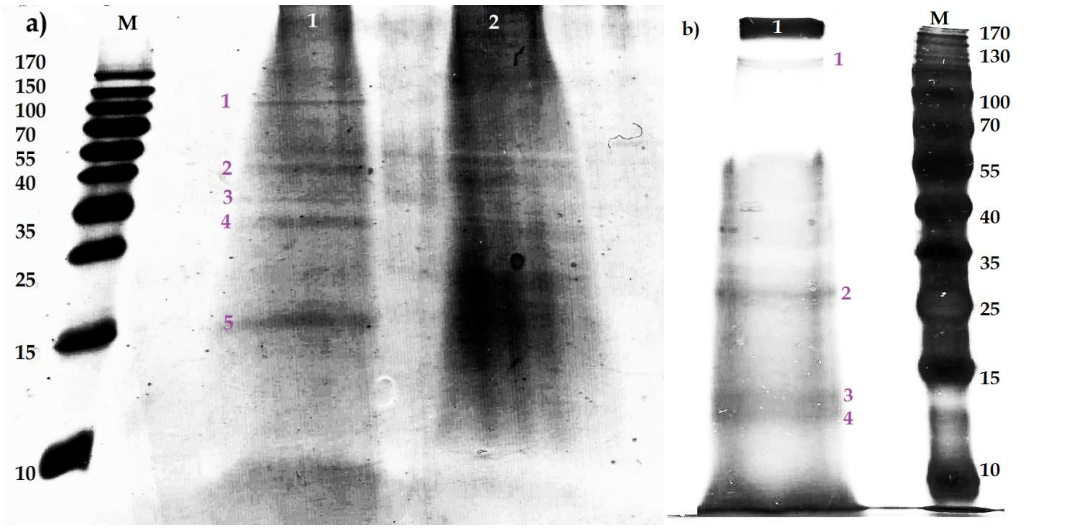


Figure 2. Purification of gamete specific glycoproteins and lectins by the use of WGA and GlcNAc affinity chromatography. A) 12% SDS-PAGE of eluted fraction collected from WGA column after loading extract from female proteins; chemically deglycosylated sample (1) and crude eluate (2). B) 10% SDS-PAGE of eluate from GlcNAc column after loading extract from male gametes. M, molecular markers with their molecular weight in kDa. Silver stained.

Considering the poor yield of proteins retained on the GlcNAc column, other affinity columns were investigated. Male and female gamete extracts were subsequently passed through mannose, fucose and fetuin columns. An absorbtion peak at 280nm was observed in the eluate from the mannose column (ca. 1mg of protein) after loading an extract from male gametes SDS-PAGE and Coomassie staining of the gel showed a strong band with a molecular mass of ca. 13 kDa (line 1 band 1, Fig.3) and three peptides corresponding to the band were identified by MALDI-TOF/TOF. Although, no match with the *Ectocarpus* protein database was found, the peptides aligned with unassigned regions on several supercontigs including sctg82 and sctg355, which are terminal contigs on the linkage group LG15 (Heesch *et al.*, 2010), and sctg716 which is unlinked. At this stage we were not able to determine the whole protein sequence and link it to the particular location in the *Ectocarpus* genome. Further studies are needed to confirm the identity and possible recognition function of this mannose specific protein. Additionally, two proteins, a catalase enzyme (Esi0043_0035) and a conserved unknown protein (Esi0052_0041), were identified in the female mannose column eluate. Only nanograms of protein were eluted form fetuin and fucose columns and could not be identified by mass spectrometry.

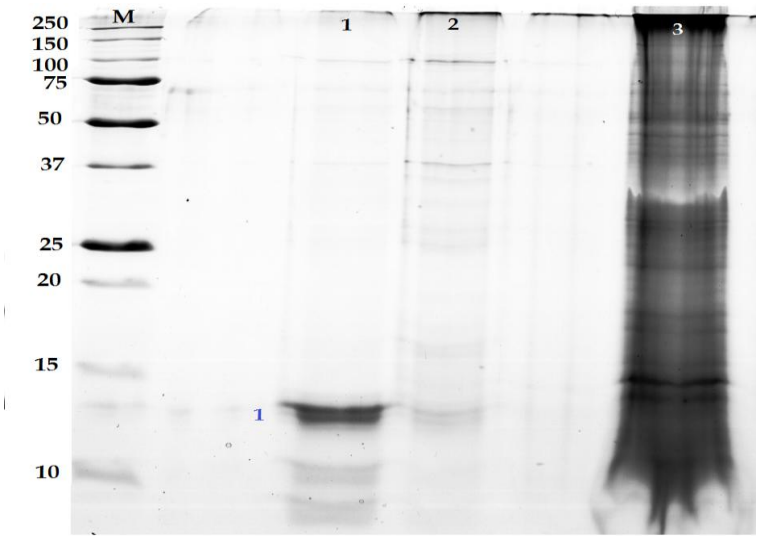


Figure 3. Purification of gamete specific lectins with mannose affinity chromatography.
15% SDS-PAGE of eluted male (1) and female (2) proteins collected from mannose column; male gametes whole cell extract (3). M, molecular markers. Coomassie staining.

3.2 Lectin induced fertilization block assay

To assess whether the gamete recognition receptors in *Ectocarpus* New Zealand strain show the same affinity as in the Port Aransas strain, we repeated the fertilization block experiment described by Schmid (Schmid, 1993). Gamete fertilization success in control samples was estimated on the level of ca. 50% as found before (Schmid, 1993). To distinguish zygotes from unfertilized gametes we looked for cells with two eye-spots, in combination with Calcofluor white staining. The latter required an additional control for the offset of parthenogenetic development.

Calcofulor white is a non-specific dye that binds to beta-linked polysaccharides such as chitin and cellulose in cell walls, and extracellular mucilage polysaccharides like ex. pectins or beta 1-3 glucans (Pawley, 2010). This fluorochrome stains most of the cell walls, but not the cell content, which made it useful in distinguishing zygotes from unfertilized cells. Unfertilized female gametes, irrespective whether they were pre-incubated with lectin solutions or not, did not show any cell wall fluorescence up to two hours from the settlement point. Only gamete excreted surface adhesives (polysaccharide-based material fixing cells to the surface) could be observed (Fig.4).

Cell-cell recognition in the New Zealand *Ectocarpus* strain could not be inhibited by any of the tested lectins, including WGA (Table 1, Fig. 5). This suggests that the male plasma membrane receptors are recognizing other carbohydrate moieties on the female cell, than in the previously investigated Port Aransas strain (Schmid, 1993). This is not surprising, since crossing experiments have shown that *E. siliculosus* represents a species complex (Stache

Crain *et al.*, 1997) and the phylogenetic analysis based on molecular markers (ITS, *rbc*L and *cox*3) separated the New Zealand strain (used in this study) from the genome sequenced strain (Peters *et al.*, 2010a). Additionally, crossing experiments indicated that *Ectocarpus* from New Zealand is distant from E*. siliculosus sensu stricto* (Stache, 1990), showing a substantial reduction of fertilization rate and zygote viability as well as the inability to undergo meiosis. This could be due to a modification of the WGA-GlcNAc system as a cause for the sterility barrier between geographically separated populations (Schmid, 1993). However no crossing experiments were done for *Ectocarpus* New Zealand and *Ectocarpus* Port Aransas strains (Stache, 1990), and our attempts to perform the experiment failed due to the inability to induce meiosis (and obtain gametophytes) in available Port Aransas parthenosporophytes.

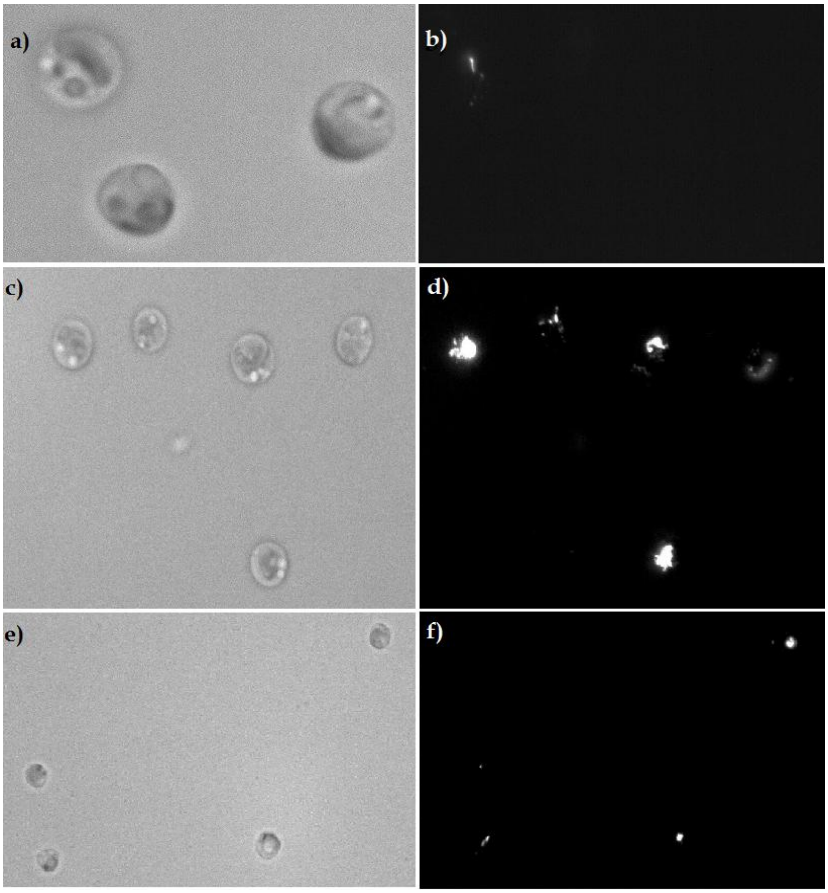


Figure 4. Settled female gametes stained with Calcofluor white.
Gametes stained 30 (a,b), 60 (c,d) and 120 (e,f) minutes after settlement. No cell wall development could be detected; white fluorescence corresponds to gamete released surface adhesives.

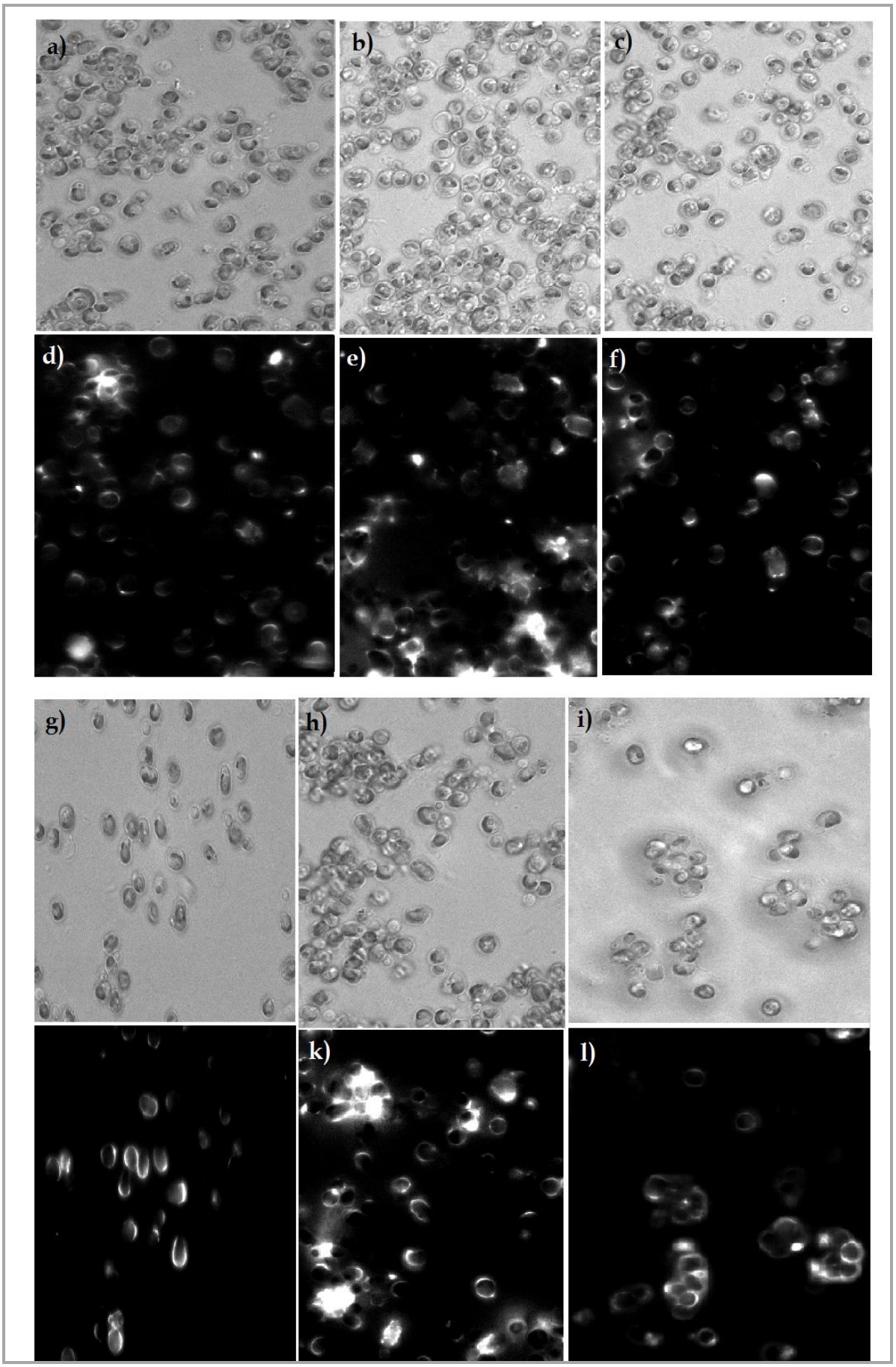

a)
b)
c)
d)
e)
f)
g)
h)
i)
k)
l)

Figure 5. Lectin fertilization block assay.
Zygotes show white fluorescence in control sample 20 minutes from gamete mixing (a,d). Fertilized female gametes pre-incubated with GNA (b,e), DSA (c,f), WGA (g,j), SBA (h,k) and PNA (i,l) lectins show no cell-cell recognition block.

Table1. Lectins used in the fertilization assay and their specificity

Lectin	Carbohydrate specificity
PNA	Gal*-β(1-3)-GalNAc
WGA	$(GlcNAcβ\ 1\text{-}4)_3$> $(GlcNAcβ\ 1\text{-}4)_2$ = GlcNAcβ 1-4Fucα 1-6GlcNAc > NeuNAc
SBA	α -GalNAc> β-GalNAc > α –Gal
GNA	Manα1-3 > Manα 1-6> Manα 1-2Man
DSA	$β(1,4)\text{-}(GlcNAc)_n$

* Man: mannose. Glc: glucose; GlcNAc: N-acetylglucosamine; GalNAc: N-acetylgalactosamine; Gal: galactose; Fuc: fucose; NeuNAc: N-acetylneuraminic acid.

4 Discussion and conclusions

We chose *Ectocarpus siliculosus* New Zealand strain (NZKU 1-3male, NZKU 32-22-21 female) to investigate gamete cell-cell recognition receptors in brown algae. It was shown that male gametes of *Ectocarpus* from different geographical locations are attracted to female gametes even if the sexual fusion is impossible (Muller, 1979a). This indicates that the chemo-attraction mechanism is far more evolutionary conserved than the sexual recognition sites for gamete fusion. Thus, rapid evolution of sex-recognition proteins may cause establishment of barriers to fertilization that, in consequence, might lead to or conclude speciation.

Sequence comparison and functional characterization of the sexual agglutinins in *Ectocarpus* could provide a valuable insight about their involvement in the speciation process. Previous studies on receptors affinity (Schmid, 1993) presented an opportunity to isolate cell-cell recognition proteins by means of affinity chromatography. However, only trace amounts of proteins were recovered using GlcNAc and WGA columns, and could not be used in *in vivo* experiments to confirm their role in fertilization. This low yield might be due to the limited number of receptors per cell, since fluorescent labeling of gamete binding sites were also unsuccessful (Maier, 1995). Nevertheless, it is also possible that the recognition proteins in *Ectocarpus* New Zealand have already substantially diverged from the *Ectocarpus* Port Aransas used in the receptor affinity study (Schmid, 1993). As for now, no gamete compatibility data for Port Aransas-New Zealand crosses are available and fertile gametophytes of Port Aransas strains could not be recovered in our laboratory.

Furthermore, we performed a fertilization block assay which did not show any inhibitory properties of WGA. Nevertheless, it is plausible that a much higher affinity of male gamete receptors to female glycoproteins than that of the plant lectin, could replace WGA at the binding site. This would reverse the fertilization inhibition effect and consequently led to zygote formation. It also cannot be excluded, that low specificity for unbranched polysaccharides used as chromatographic ligands in this study, or on the contrary - high specificity of the lectin and affinity column interaction -, hindered purification results. We were, however, able to purify a mannose-binding protein from male gametes, whose involvement in fertilization still has to be confirmed.
Future studies using lectins with a broader affinity spectrum (including algal agglutinins) as well as experiments using gametes of other *Ectocarpus* strains (Port Aransas strain in particular) may bring more conclusive results.

Because the biochemical isolation of sex-recognition proteins caused difficulties, the whole transcriptomes of male and female gametes were sequenced to explore their functional differentiation at the molecular level (see the following chapters).

Chapter 3

Generation of transcriptome data using Next Generation Sequencing technology

A.L. contributed to this chapter by collecting gamete material, performing RNA extractions and RNA-Seq data analysis.

1 Introduction

Recent technological advances in DNA sequencing have opened a new era in genomics and fundamentally changed our ability to identify genes and understand gene expression, by making comprehensive genome analysis accessible and relatively inexpensive (Mardis, 2008)(Koboldt *et al.*, 2013). Since the development of automated Sanger sequencing (Sanger *et al.*, 1977) in the 1980s, sequencing technology has shifted from chromatographic gels, through capillary electrophoresis, to state-of-the-art, high-throughput next-generation sequencing (NGS) and even beyond, to the Single Molecule Real Time sequencing (SMRT). NGS generates hundreds of thousands to tens of millions reads with individual lengths varying between 30 and 800 bp. This advancement in coverage and sequencing depth allows biologists to address many questions in transcriptomics, including de novo gene discovery, alternative splicing, small RNAs and chromatin modification that impact gene expression (Simon *et al.*, 2009)(Werner, 2010). NGS methods allow to quantitatively assess the transcriptional levels based on the reads matched to a reference sequence and enable comparative expression studies, competing with microarrays in this field. Since microarrays suffer from background noise, cross-hybridization of RNA probes and are limited only to genes represented on the array, NGS appears to be more suited for RNA discovery and abundance measurements. Several NGS platforms are commercially available including: 454 Pyrosequencing (Roche); Illumina's sequencing by synthesis (SBS) and sequencing by oligo ligation and detection (SOLiD) from Applied Biosystems (for overview of the platforms see (Voelkerding *et al.*, 2009)). Considering the availability of the *Ectocarpus* genome sequence, SOLiD v.3 sequencing was chosen for the purpose of this study to generate short, randomly distributed reads and gain insight into the functional complexity of the *Ectocarpus siliculosus* gamete's transcriptome. Here we present the analysis and characteristics of the RNA-Seq Next Generation Sequencing data generated for *Ectocarpus* gametes. Quality evaluation, different stringency alignment modes and possible drawbacks are discussed, preparing ground for functional analysis of the expressed genes in the following chapter.

1.1 SOLiD sequencing characteristics

The SOLiD technology platform uses emulsion PCR and sequencing by oligonucleotide ligation and detection rather than sequencing by synthesis (http://solid.appliedbiosystems.com). A library of DNA fragments is clonally amplified with SOLiD adapters onto polystyrene beads by means of emulsion PCR. The beads are subsequently enriched and covalently bound to the glass slide divided into one, four or eight segments. In the following step, sequencing primers complementary to the SOLiD adapters

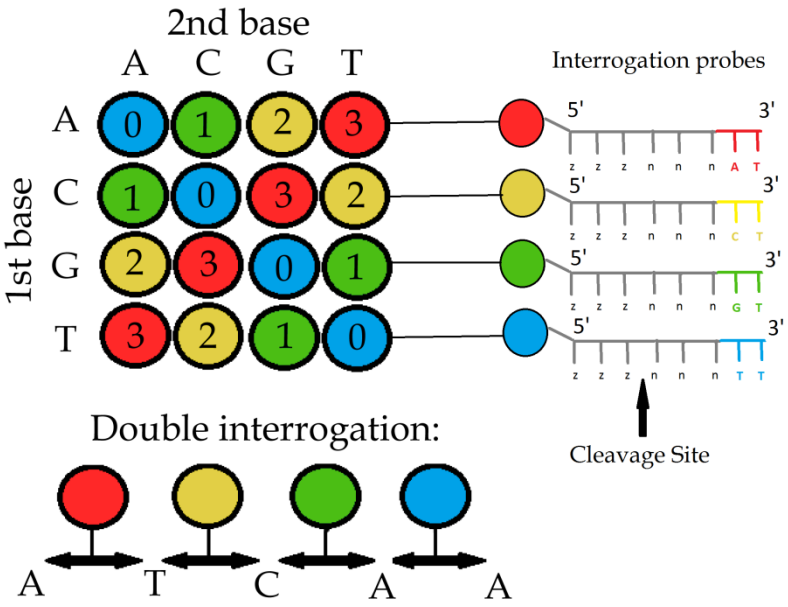
and a mixture of four different fluorescent oligonucleotide probes are added, which compete for ligation to the primer. The SOLiD unique sequencing methodology involves ligation of probes that cover two base positions at a time, which implies that four different dyes code for sixteen possible two-base configurations (Fig.1a). The color code scheme is based on the Klein four-group, which is a symmetry group of a rectangle. The unique properties of this design permit for each di-base and its reverse, complement, or reversed complement to have the same color, but restricts two di-bases with the same first base to have different colors. This system provides tools to discriminate between measurements errors and true sequence variation (ex. SNPs).

Multiple cycles of ligation, detection and cleavage are performed until the desired read length is reached. The extension product is than removed and the process is repeated using a new primer shifting the start position of the sequencing by 1 base. Five rounds of primer reset are completed for each sequence tag, which allows interrogating each nucleotide two times independently (Fig.1b). The overlap of nucleotide dimers and the nature of the color code facilitates error correction and incorporates an extra quality check. Nevertheless, this system creates an obstacle in generating strings of color spaces rather than nucleotides, which requires the knowledge of the first sequenced base to correctly translate to nucleotides. Therefore, the first base of the tag, which originates from the SOLiD adaptor, is always provided with the color coded sequence. Nevertheless, sequencing errors may introduce frame shifts inside the sequence which hinder the proper decoding. To minimize the likelihood of this type of errors, all of the downstream analyses are carried out in color space, and this may require conversion of the reference sequence (ex. genome sequence) to color code prior to alignment of the reads.

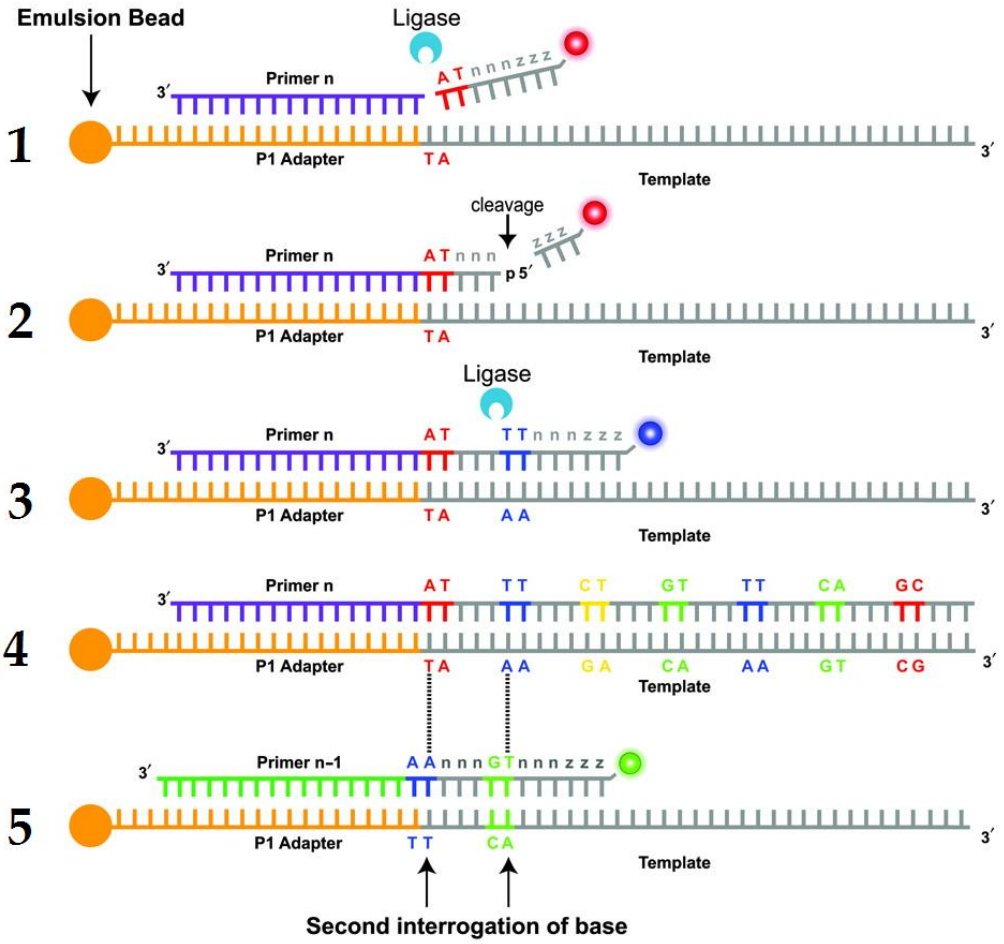
1.2 SOLiD Whole-Transcriptome sequencing

NGS has provided a powerful tool for measuring and quantifying transcriptome composition, namely a whole-transcriptome or RNA-Seq approach. A typical protocol starts with poly-A selected mRNA or total, rRNA depleted RNA sample and involves shearing, adaptor ligation, conversion to cDNA and amplification; although these steps may appear in a different order depending on the sequencing platform used. SOLiD sequencing depends on RNA shearing prior to attachment of 5’ and 3’ specific adapters and PCR amplification, which allows retaining the strand specificity in the output. The size selection step on the fragmented RNA is influenced by the desired data output. Longer pieces are more suitable for obtaining splice variants information or de novo transcript assembly, whereas shorter sequences are commonly used in quantification studies.

a)


2nd base
A C G T
1st base
A 0 1 2 3
C 1 0 3 2
G 2 3 0 1
T 3 2 1 0
Interrogation probes
5' 3'
z z z n n n A T
5' 3'
z z z n n n C T
5' 3'
z z z n n n G T
5' 3'
z z z n n n T T
Cleavage Site
Double interrogation:
A T C A A

b)


Emulsion Bead
Ligase
1
Primer n
A T n n n z z z
3′
P1 Adapter
T A
Template
3′
cleavage
2
Primer n
A T n n n
z z z
p 5′
P1 Adapter
T A
Template
3′
Ligase
3
Primer n
A T
T T n n n z z z
3′
P1 Adapter
T A
A A
Template
3′
4
Primer n
A T T T C T G T T T C A G C
3′
P1 Adapter
T A A A G A C A A A G T C G
Template
3′
5
Primer n–1
A A n n n G T n n n z z z
3′
P1 Adapter
T T
C A
Template
3′
Second interrogation of base

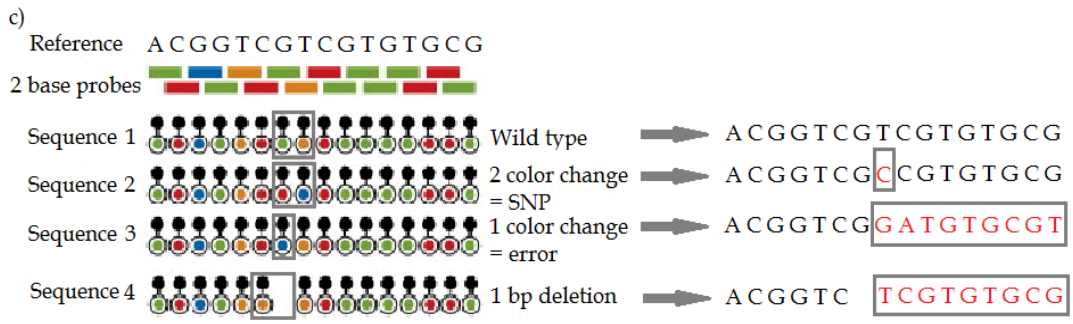


Figure 1. a) SOLiD Color Space Coding scheme. Each probe is an octamer build from 2-probe specific bases followed by 6 degenerate bases (nnnzzz) and a fluorescent probe linked at the 5' end. The fluorescent signal is read after annealing and ligation of the probe, before the cleavage of the last 3 degenerate bases. b) Schematic representation of SOLiD sequencing by ligation. (1) The first round of interrogation, in this example the first sequences bases are AT; (2) Cleavage of the last 3 bases and phosphorylation of the 5' end; (3) Annealing and ligation of the second probe; (4) Complete product of the first round of interrogation after 7 annealing cycles; (5) Second round of interrogation. Annealing of the new primer of a length (n-1) shifts the sequencing start position and allows second time interrogation of the sequenced bases. c) The principle of 2 base encoding showing the method for distinction between sequencing errors and true polymorphism or indels. Note that one sequencing error causes the remaining sequence mistranslation if changed directly to nucleotides. Modified from (Voelkerding *et al.*, 2009).

Counting read density corresponding to genes or exons provides comprehensive information about transcript quantity and improves gene models by assigning UTRs, splice sites and new transcriptionally active sites. The latter gives RNA-Seq a big advantage over micro-arrays which are constrained to existing genomic sequences, making it impossible to characterize a transcript without prior knowledge of thegenomic origin. RNA-Seq is also capable of single-base resolution which gives it a greater ability to distinguish RNA isoforms, determine allelic expression, and reveal sequence variants. The broad dynamic range of RNA-Seq improves performance for quantitative detection of low and highly abundant transcripts.

2 *Materials and methods*

To explore transcriptional differentiation between male and female gametes of *Ectocarpus,* total RNA was isolated from freshly released, flash-frozen gametes (ca. $4 \cdot 10^6$-$4 \cdot 10^7$ cells per sample) using the Machery-Nagel NucleoSpin RNA XS kit according to the manufacturer's instructions. Initial RNA quality measurements (NanoDrop readings and agarose gel) can be found in the Supplemental Figure 1. Total RNA (20µg per library) was depleted of ribosomal

RNA and fragmented using RNase III. In the following step the 3' and 5' specific adapters were added, RNA fragments were reverse transcribed to cDNA which was then selected for a size range of 160-280bp and amplified using PCR. Barcodes specific to male and female sample were introduced with the 3' PCR primer during amplification, which later allowed combining the samples for sequencing on 1/8 of a plate. Only the sense-strand of the RNA was sequenced.

RNA sequencing and library preparation were performed at Cofactor Genomics (Missouri, USA). The schematic representation of the SOLiD Whole Transcriptome sequencing workflow is shown in Figure 2.

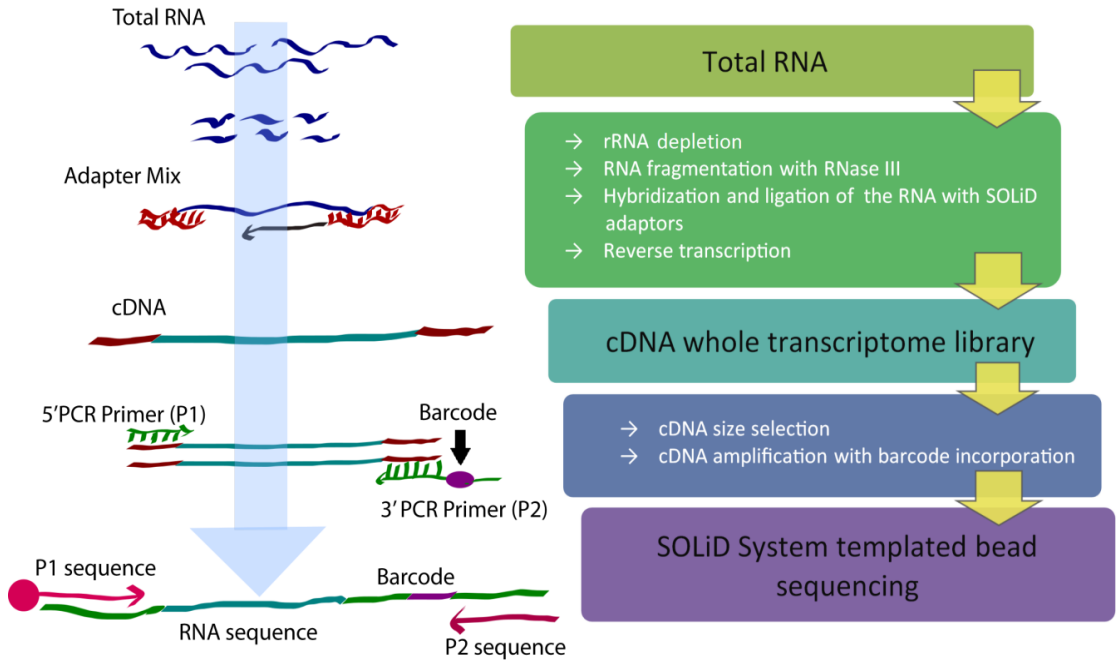


Figure 2. SOLiD Whole Transcriptome analysis workflow.

3 *Results*

3.1 *Initial data processing and assessment of reads quality*

More than 65 million, 50bp AB SOLiD reads, were generated during a single run. Initial processing included adaptor pruning and filtering of reads shorter than 50bp. Raw sequence data composed of 36 million reads for the male sample and 28 million reads for the female sample were obtained (color space sequence FASTA files (CSFASTA) and matching quality files (CSFASTQ)). The quality values (QVs) in the SOLiD system are in linear relationship with

predicted and observed phred-scale quality scores (number assigned to each sequenced nucleotide in automated sequencer traces (Ewing *et al.*, 1998)) and are calculated based on training the sequencing process parameters against multiple annotated datasets. In general, QVs represent the confidence in color call accuracy and the score Q is defined by Q = -10 log10(P); for example, Q = 30 corresponds to the probability P = 0.001 that a base has been called incorrectly. An example of a color coded read and its corresponding quality string is shown below (color calls are represented by 0-3 digits).

```
>949_910_270_F3
T000101102030111100311311002203210212210111011000 10
>949_910_270_F3
30 11 27 20 31 30 32 23 27 12 23 26 22 28 28 29 29 17 18 31 24 30 20 24 21 30 25 10 5
28 31 30 5 17 14 31 22 13 6 30 15 30 6 7 27 13 29 12 12 22
```

Since Chaisson (Chaisson *et al.*, 2009) reported that the number of errors in SOLiD reads, similarly to other NGS platforms, increases towards the 3' end and SOLiD platform reports all sequenced reads after the primary analysis without quality pre-filtering, it is possible to encounter poor-quality sequences with miscalls (random single base error) or polyclonal calls (missequencing due to polyclonal beads). It is therefore essential to perform a statistical quality assessment prior to processing the reads, since the overall quality will influence the error tolerance settings in downstream analyses.

The quality statistics for male and female gametes, obtained with the FASTQ statistics tool (Blankenberg *et al.*, 2010) implemented in the Galaxy resource (https://main.g2.bx.psu.edu/) (Blankenberg *et al.*, 2001)(Giardine *et al.*, 2005)(Goecks *et al.*, 2010) are shown in Fig. 3. The overall read quality was good with very high median scores (QV ≥ 25) for the first 30 bp. Nevertheless the characteristic quality drop towards the 3' end was visible in both samples. For this reason as well as the fact that our sequencing data came from a different *Ectocarpus* strain than the one used for genome sequencing, we applied more relaxed, error tolerant settings for alignment to the reference and selected for best matching reads based on the mapping quality parameters.

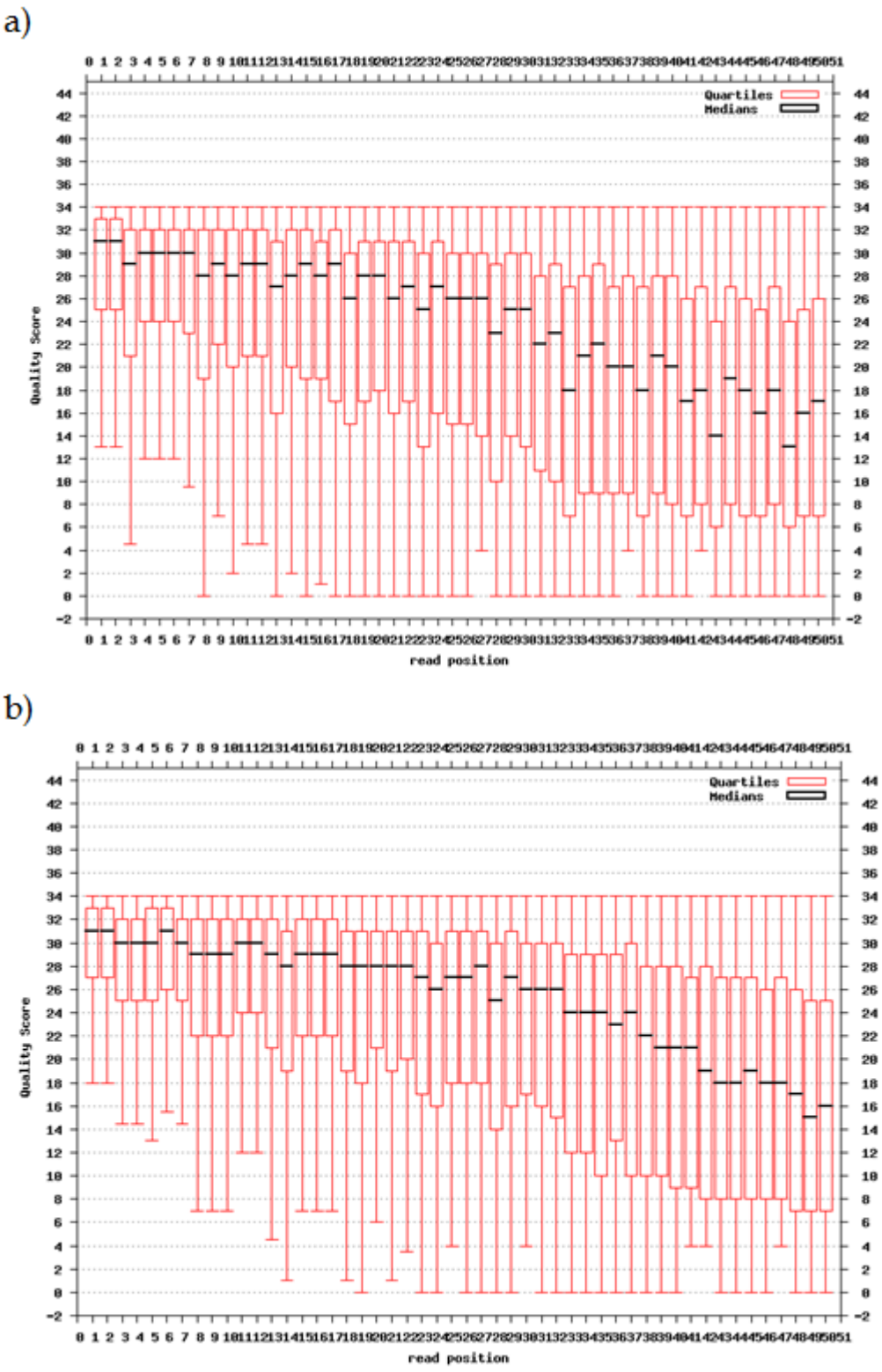


Figure 3. Quality score distribution depending on the base position for female (a) and male (b) sequence reads.

3.2 Sequence alignment and pruning of sub-optimal matches

SHRiMP2 (the SHort Read Mapping Package) (Rumble *et al.*, 2009) was chosen to map the reads to the reference sequence due to its ability to handle data with sequence polymorphisms. SHRiMP is a set of algorithms and methods designed specifically to map short, color-space reads to a genome and a statistical model called ProbCalc for distinguishing false positive hits. Using the SHRiMP algorithms each read is matched against the whole genome sequence and several matching positions with variable polymorphism are reported. In the next step, the most probable variants are selected based on the statistical scoring of the hits and false positives are filtered out (Fig. 4).

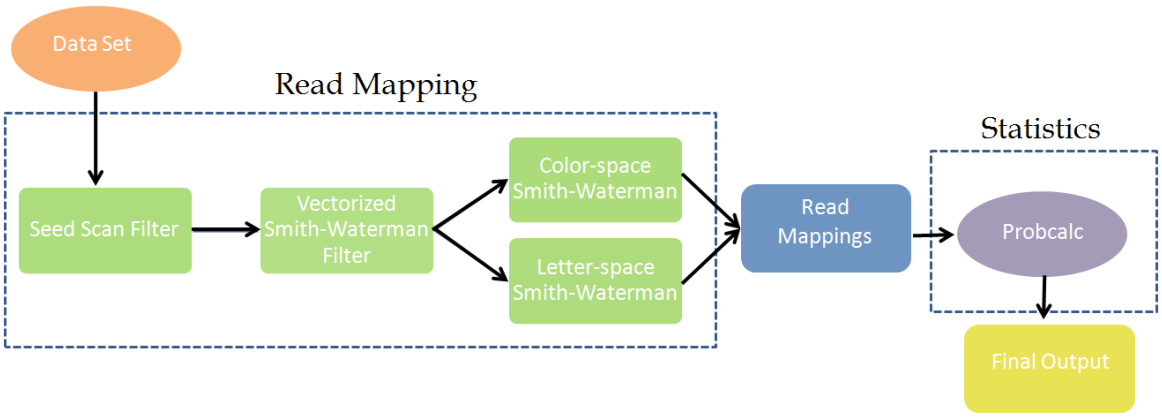


Figure 4. Data flow and processing with SHRiMP. Candidate mapping positions are first discovered with a seed scanner and later validated with the Smith-Waterman algorithm. Top scoring hits are fully aligned and the statistical significance is calculated.
From: Rumble SM, Lacroute P, Dalca AV, Fiume M, Sidow A, et al. (2009) SHRiMP: Accurate Mapping of Short Color-space Reads. PLoS Comput Biol 5(5).

To map *Ectocarpus* sequences we set the threshold score for full Smith-Waterman (SW) alignment to 60% of the maximum perfect match score to permit for more sequence dissimilarity. The Smith-Waterman algorithm allows typical parameters, including mismatch, indels and specifically implemented 'crossover' penalties. The latter is calculated for reads in which a color-space error causes the rest of the sequence to be mistranslated (see Fig. 1c, Sequence 3), but the genome will match one of the other three possible translations performed simultaneously by the algorithm. This feature of SHRiMP permits mapping of the SOLiD reads directly to the nucleotide genome sequence, without translating it to color codes. The detailed penalty scores used in the alignment are listed below.

SW Match Score:	10
SW Mismatch Score:	-15
SW Gap Open Score (Ref):	-40
SW Gap Open Score (Qry):	-40
SW Gap Extend Score (Ref):	-7
SW Gap Extend Score (Qry):	-7
SW Crossover Score:	-14

Statistical significance of top scoring hits was assessed using the Probcalc module of SHRiMP with final mapping output files and genome size as a parameter. Probcalc computes the probability that a hit would be generated by chance in a random genome ('pchance'), the probability that it is generated by the reference genome ('pgenome'), and the normalized odds ('normodds').

'pchance'	-	Probability that a read will align with a genome with as good a score or better by chance.
'pgenome'	-	Probability that a hit was generated via common evolutionary events characteristic of the genome.
'normodds'	-	Normalized pgenome/pchance.

The first value ('pchance') is given based on the probability that a matching read would map equally good or better in the genome of similar length where at every position each nucleotide can be randomly selected with a probability of 0.25. The second value ('pgenome') evaluates the probability that a certain region of the genome generated a given read, based on the frequency of matches, mismatches, indels and crossovers observed in the alignment.

For example, a good alignment would have a low 'pchance' (close to 0) and a very high 'pgenome' (close to 1). Next the normalized odds ('normodds') are computed for every hit by every read by summing up the odds ('pgenome'/'pchance') and dividing each value by their sum. The output value allows estimating if the hit is more likely to be the right one comparing to other ones. This approach creates large discrepancies between having an almost exact match ('normodds' close to 1) and a more distant one ('normodds' close to 0). Two equally good hits will have normalized odd values of 0.5 for each of them and one single hit would have a 'normodds' value of 1.

Despite development of many algorithms and tools for mapping reads to the reference genomes, accurately mapping RNA-seq reads remains a tough problem due to the complexity of the transcriptome. To examine sequencing alignment data according to the features of the mapped reads and their genomic regions, we applied stringent ('normodds'≥0.7) and more relaxed ('pchance'≤0.3) filtering.

Reads having multiple hits with equal SW scores were discarded and only unique alignments were considered for further analysis. Considering that the *Ectocarpus* genome is rich in introns, with on average seven introns per gene (average size 704bp), longest introns reaching 20,000bp and 99.6% of all introns being equal or shorter then 5000bp, we used the Tophat software (Trapnell *et al.*, 2009) to identify reads mapped in exon-exon splice junctions. Taking into account intron characteristics, we allowed for 1 mismatch and an intron length of maximum 5000 bp.

3.3 Genome correspondence depending on filtering criteria

We compared numbers of reads mapped to nuclear, chloroplast and mitochondrial genomes depending on selected filtering thresholds, the nuclear mappings were also divided into annotated genes and intergenic regions. The results shown in Figure 5 indicate a substantial increase in read mapping efficiency when less strict filtering conditions were applied. 45% of total male reads and 62% of total female reads were aligned to the reference genome with 'pchance' ≤ 0.3 comparing to only 9% for male and 13% for female if 'normodds' ≥0.7 was a criterion. However, the majority of the additional mappings in the first scenario fall into the chloroplast genome or intergenic regions of the nuclear genome. Closer evaluation of these reads indicated mappings in repetitive sequences, which are otherwise removed from the final output in the second scenario, since 'normodds' values take into account the characteristics of the genomic region ('pgenome'). In the next step, we used HTSeq-count to locate and count aligned reads within annotated genes, based on the available *Ectocarpus* gene set (Sterck *et al.*, 2012). With the HTSeq-feature count we also determined the number of reads mapped in intronic, 3' UTR and 5' UTR regions (Fig. 6). We observed an evident increase in intron mappings when the filtering conditions were less stringent, altogether the intronic part accounted for as much as 51% and 73% of gene mappings in male and female respectively. When the two filtering conditions were compared for numbers of detected genes, there was no significant difference found. Different results were obtained, however, when a threshold of at least 5 mapped reads covering 51 bp of an exon was set as criterium for genes being considered expressed and the 'pchance' data set contained ca. 5% more genes than the 'normodds' set. To investigate the influence of the filtering method on gene expression levels, we compared the number of mapped reads for each gene (only coding

sequences were considered) for the two filtering conditions. Scatter plots in Figure 7 shows that with more stringent filtering we do not change the landscape of relative gene expression levels in a cell, which is represented by the proportional relationship between numbers of mapped reads in each data set. To minimize the possibility of false positives related to repetitive sequences we focused on a high quality subset of the data consisting of sequence variants supported by 'normodds' values ≥ 0.7 for the actual analysis (Table 1).

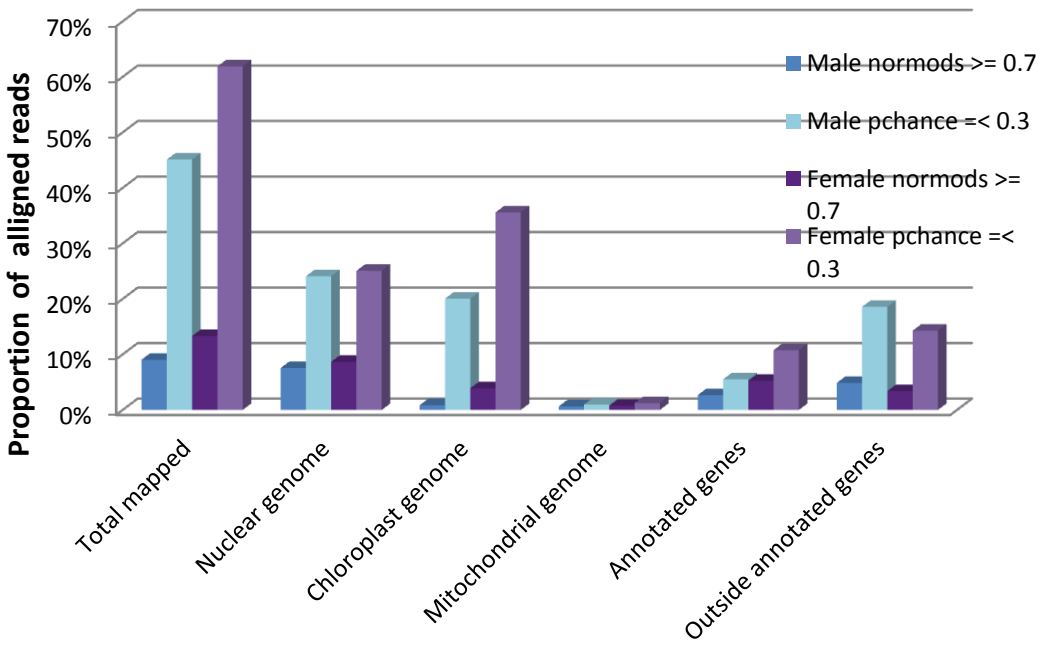


Figure 5. The influence of filtering parameters on the proportion of reads (scaled to the total number of sequenced reads) mapped to the reference. To examine sequencing alignment data we applied stringent ('normodds'≥0.7) and more relaxed ('pchance'≤0.3) filtering parameters for male and female samples; corresponding reference sequences included nuclear, chloroplast and mitochondrial genomes as well as annotated genes and intergenic regions.

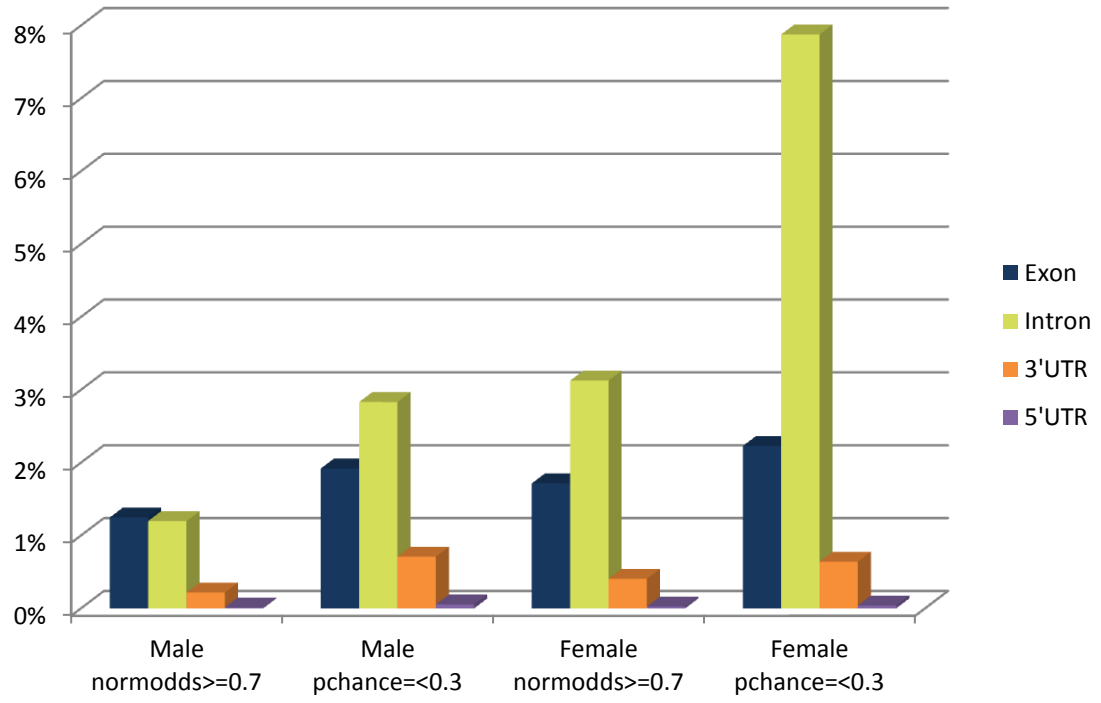


Figure 6. Proportion of reads (scaled to the total number of sequenced reads) counted in gene regions (exon, intron, 5' UTR, 3'UTR) for male and female samples with 'normodds'>=0.7 and 'pchance'=<0.3 filtering parameters.

Table 1

Transcriptome mapping results of *Ectocarpus siliculosus* male and female gametes

	Male gametes	Female gametes
Total number of reads	36 751 768	28 591 842
Total number of mapped reads	16 580 350	17 697 894
High quality mappings (normodds>=0.7)	2 775 695	2 478 987
High quality mappings in known genes	899 347	1 386 445
High quality mappings outside of known genes	1 876 348	1 092 542

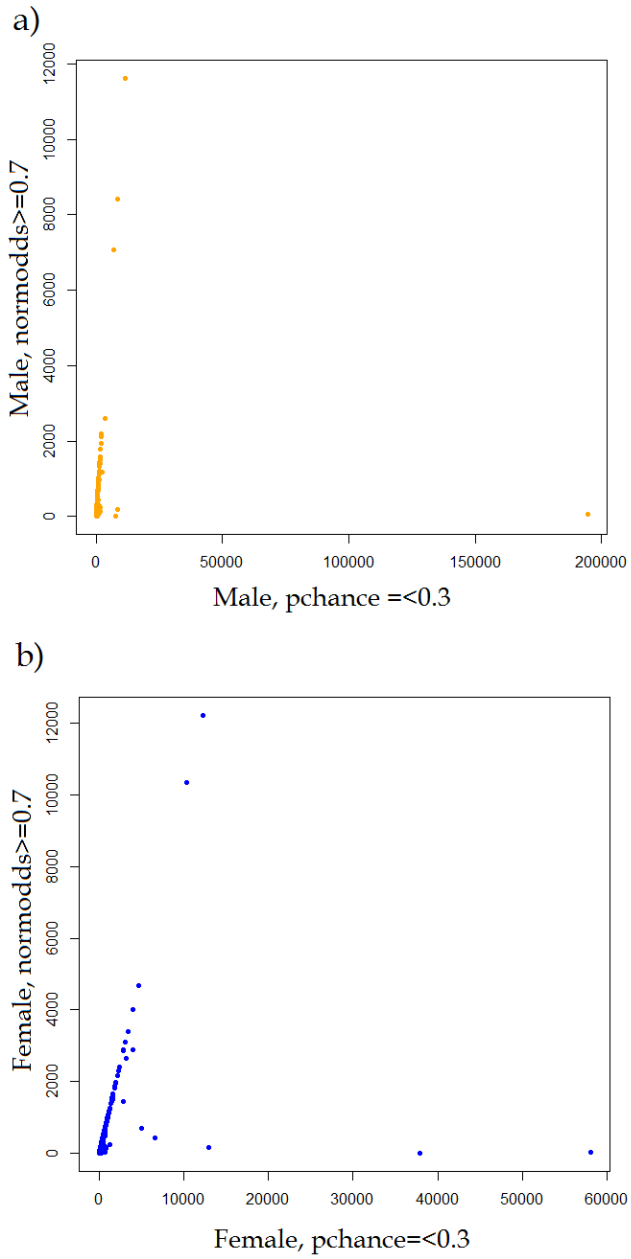


Figure 7. Scatter plot showing how gene expression levels vary in proportion to the number of mapped reads, depending on the chosen filtering method.

3.4 High quality mappings

Following alignment quality trimming with 'normodds'≥ 0.7 cutoff value, 16.74% of all aligned reads for male and 14.01% of the reads for female were chosen as unique and high quality mappings to the nuclear genome (Table 1). Nevertheless, more than 80% of the alignments fell outside the coding exons. As it is illustrated in Figure 6, part of them comes from intronic regions which represented 44% and 59% gene mappings for male and female

respectively. In general rRNA depleted samples are richer in intron mappings compared to poly-A selected samples. However, it is not clear whether these intronic reads originate from independent transcripts located within introns, or whether they represent immature transcripts that had not been spliced (Kapranov *et al.*, 2010)(Wetterbom *et al.*, 2010). Recent studies in humans indicated that nascent transcription in combination with co-transcriptional splicing are widely occurring, resulting in as much as 75% of whole RNA transcripts in brain tissue with mappings outside of known exons (Ameur *et al.*, 2011). These findings support the high proportion of reads mapped to introns from pre-mRNA. Nevertheless, it is not known if and to what extent co-transcriptional splicing occurs in *Ectocarpus*. Thus, intron aligned reads were omitted in the consecutive analysis.

The remaining portion of reads that did not match annotated genes suggests unannotated 5' and 3' UTRs or other transcriptionally active sites. These findings are in line with the whole genome tiling array, which identified 8,741 expressed regions longer than 200 bp outside of predicted genes as potential novel protein-coding regions or non-coding RNA genes (Cock *et al.*, 2010).

Raw data were also aligned against the mitochondrial and chloroplast genome of *Ectocarpus*, using the same filtering parameters. Around 0.86% and 3.90% of total reads mapped uniquely to chloroplast genome and 0.64% and 0.76% were aligned to mitochondrial genome (male and female respectively). These low percentages might be due to the short length of organelle genes, which would be lost from the sequencing pool after fragmentation and size selection. Secondly, organelle genomes account only for a small proportion of coding sequences as a result of gene transfer to the nucleus (mitochondrial genome in *Ectocarpus* encodes 62 genes, whereas 605 mitochondrial targeted genes are found in the nuclear genome) (Cock *et al.*, 2010). Only exon mapped reads were considered in further analysis.

3.5 Gene expression analysis

A gene was considered expressed when the minimum depth of coverage requirements were met, i.e. at least five reads covering minimum 51bp of the coding sequence. Applying these criteria we found 8,029 and 7,777 out of 16,239 annotated genes expressed in male and female gametes respectively. The transcription level of each gene was determined by the number of reads mapped in the exon region normalized to gene length (kbp) and total reads number in millions (RPKM). Genes were classified into four expression categories: 'Very highly expressed' (RPKM >= 100), 'Highly expressed' (100 > RPKM >= 10), 'Medium expressed' (10 > RPKM >= 1) and 'Low expressed' (1 > RPKM). As shown in (Fig. 8) the sequencing data are enriched in medium and low expressed transcripts, confirming that RNA-

Seq is sensitive for lowly expressed genes that would otherwise be undetected (Werner, 2010).

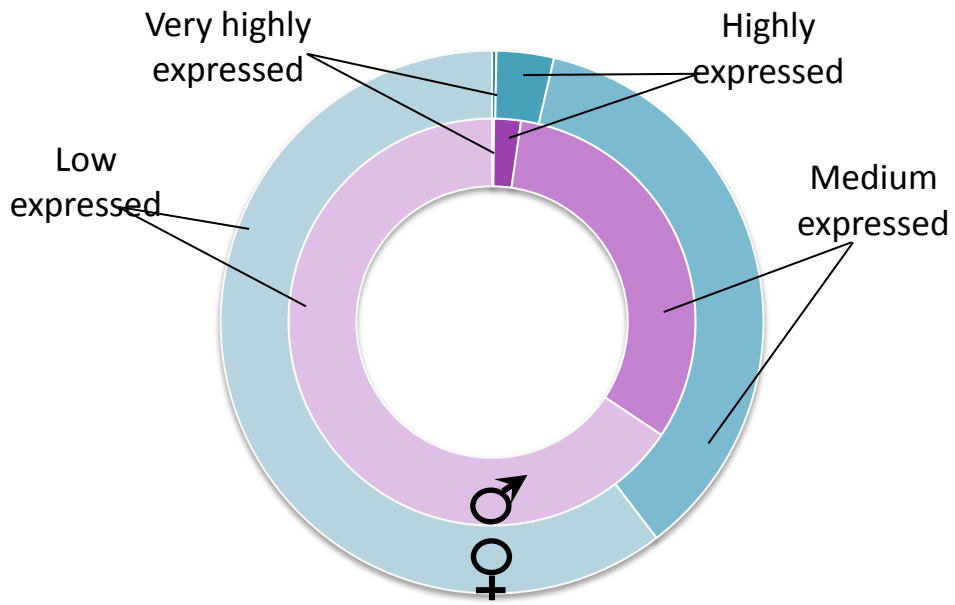


Figure 8. Overview of gene classification into expression categories based on mapped reads number normalized to gene length (kbp) and total reads number in millions (RPKM). Genes were grouped into four categories: 'Very highly expressed' (RPKM >= 100), 'Highly expressed' (100 > RPKM >= 10), 'Medium expressed' (10 > RPKM >= 1) and 'Low expressed' (1 > RPKM); male gametes (purple circle) and female gametes (blue circle).

3.5.1 Differential gene expression

In order to obtain biologically meaningful information about gene expression, it is necessary to normalize libraries. When some methods calculate values in units of reads per kilobase per million mapped reads (RPKM), accounting for library size seems to be more intuitive for a few reasons. First of all, considering that each sequencing method has a certain maximum depth, sequencing sample to half its depth will render half the reads mapped to each gene. Additionally if one sample is rich in transcripts that are unique and highly expressed, the sequencing capacity for other genes decreases. The hypothetical examples above highlight the fact that composition of the RNA pool, rather than the gene length, is a crucial factor to estimate gene expression. Therefore, we applied a method that uses raw read counts and library sizes to estimate scaling factors for downstream statistical analysis. The method estimates the ratio of RNA production per sample with the weight trimmed mean of the log expression values (trimmed mean of M-values method (TMM) (Robinson & Oshlack, 2010)).

The obtained normalization factors were 1.10 for the male and 0.91 for the female library. The TMM normalization method is implemented in edgeR package for R which was used to perform differential expression analysis.

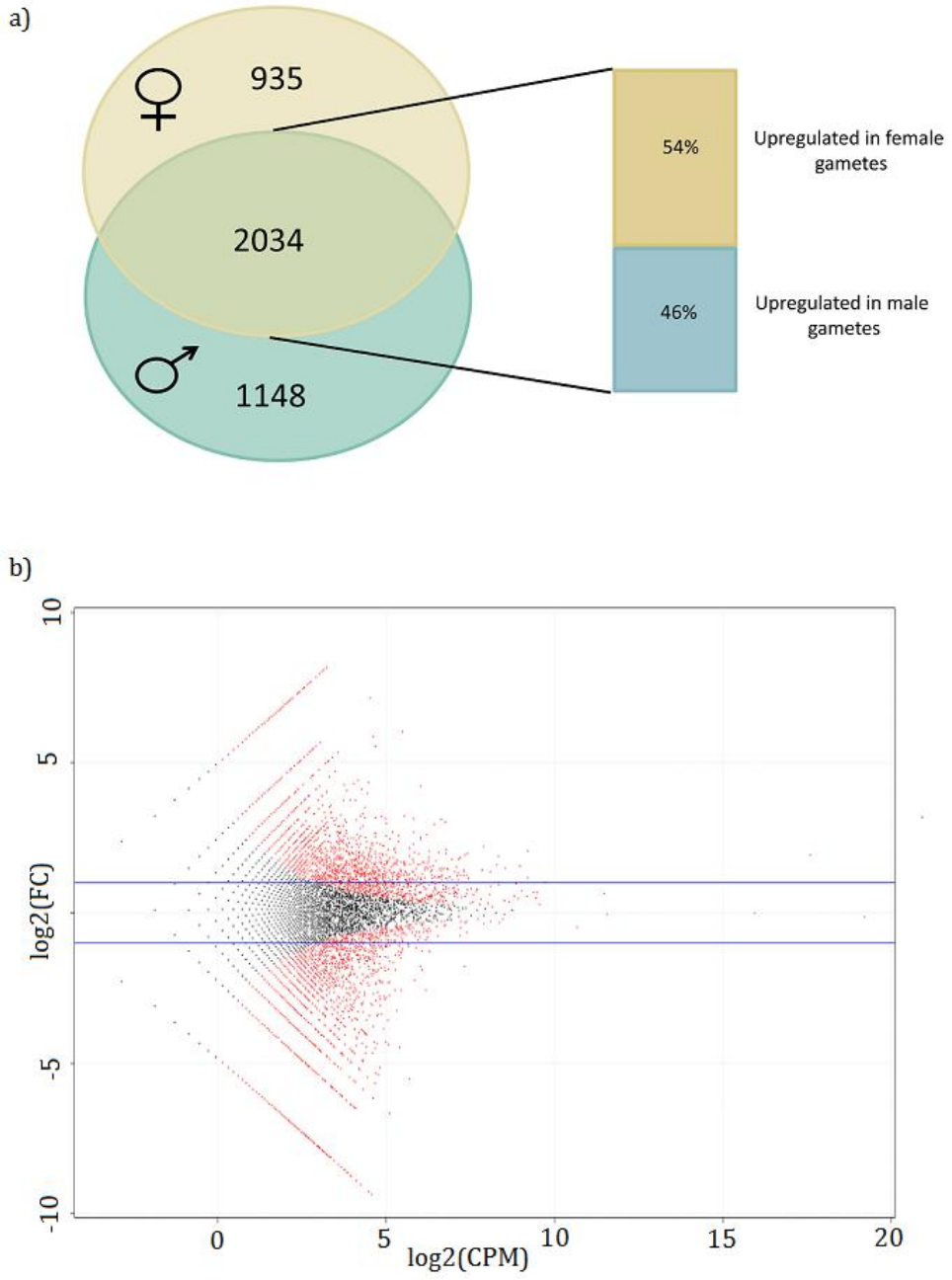


Figure 9. Differential gene expression between male and female gametes of *Ectocarpus*.
a) The distribution of 4117 DE genes with Fold Change >= 2 at FDR < 5%. b) Scatter plot showing distribution of fold-change in expression in male versus female gametes (y-axis) against expression level (x-axis). Differentially expressed genes at significance level of FDR < 5% are colored red; blue line indicates 2-fold difference in expression.

We identified 4117 differentially expressed (DE) genes between male and female gametes allowing a false discovery rate (FDR) of 5% and a fold change >= 2. As shown in Figure 9a, 1148 of these genes were expressed only in male and 935 genes were found only in female gametes. 54% of the remaining 2034 genes, shared between the two samples, were upregulated in females. Figure 9b displays the distribution of fold change (log2(FC)) as a function of gene abundance level represented as log2(CPM).

Since edgeR uses raw read counts and library size as input to calculate normalization values (TMM method) and differential expression, we tested the DE data for biased coverage, which would favor the long transcripts or genes with a large number of mapped reads. As shown in Figure 10, there is no correlation between gene length and fold change (a) or between number of mapped genes and detected fold change (b).

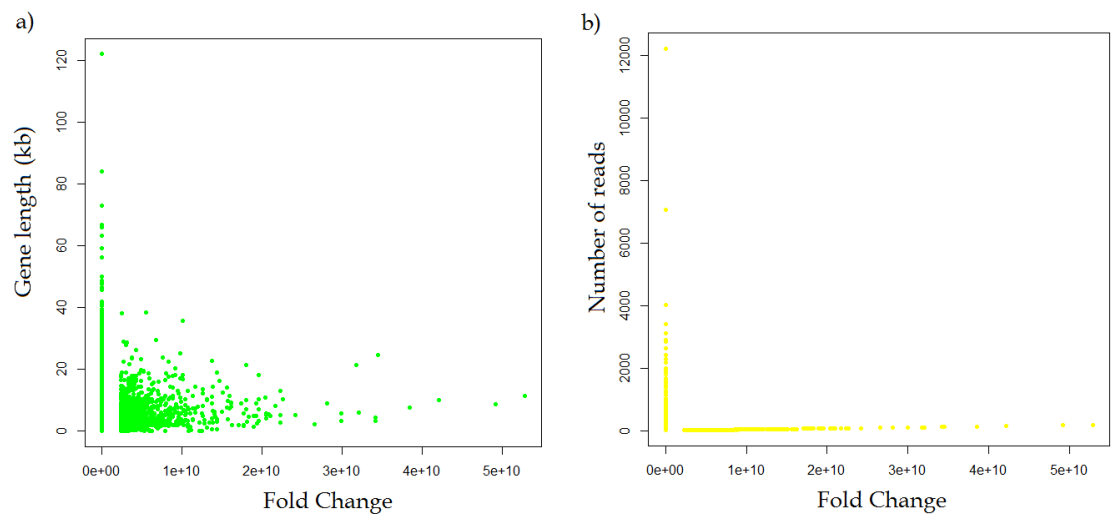


Figure 10. Test of differential expression bias. a) Detected Fold Change (FC) as a function of gene length (kb). b) Detected Fold Change as a function of the number of mapped reads.

4 *Discussion and Conclusions*

Next generation sequencing technologies can be used to characterize and screen transcriptomes at great depth and coverage. However, a number of significant challenges need to be addressed in gene expression analyses based on short-read NGS data. We aligned the transcriptome of male and female gametes of *Ectocarpus siliculosus* New Zealand strain (NZKU 1-3 male and NZKU 32-22-21 female) to the reference genome. We found that the degree of differentiation has to be taken into account while establishing the alignment parameters even when sequencing closely related organisms. Although it is now straightforward to set up RNA-Seq experiments, data interpretation and understanding the

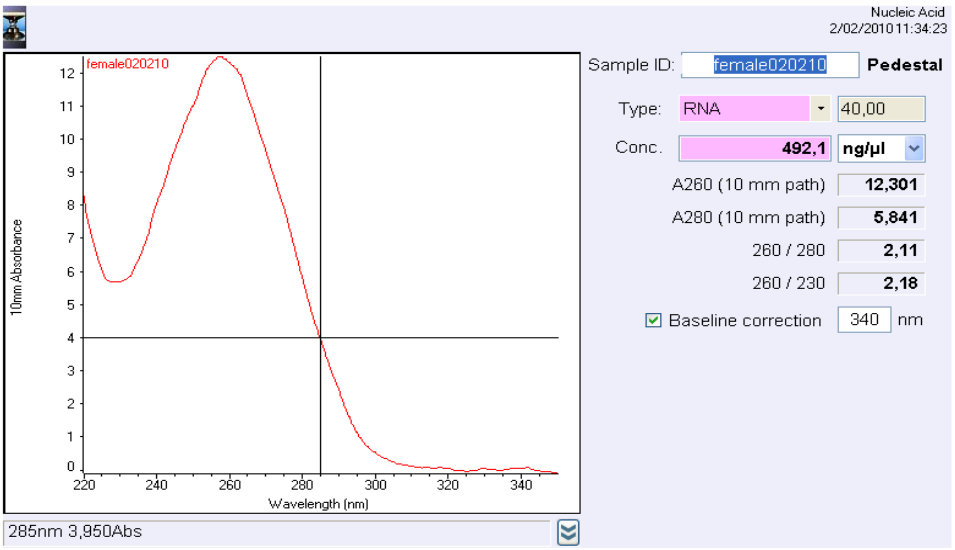
intricate transcriptional architecture still fall behind. Ribosomal RNA depleted RNA-Seq samples generate a static snapshot of a dynamic transcriptional state in a cell, with RNA transcription, translation, silencing and turnover in the progress. In this instance a deduction of the true status of transcripts and their role in cell functioning is a challenging task. Important aspects of dealing with the RNA-Seq data are the processing methods used to extract meaningful biological information. Deciding on the gene expression level also pose difficulties, since many reads align in intronic regions and it is not clear if they represent pre-mRNA or independent transcripts. Using only exon mapped reads does not fully solve the problem, particularly if intron mapping represents a substantial part of the alignment, like in the case of *Ectocarpus* gametes, because it may lead to underestimation of gene expression. It should be noted, that intronic regions represent as much as 40% of the *Ectocarpus* genome sequence, whereas exons constitute only 16% (Cock *et al.*, 2010). If intron mappings belong to pre-mRNA transcripts in *Ectocarpus* gametes, we can expect they would determine a much higher proportion of reads compared to exon mappings. Additionally, the complexity of a transcriptome itself hinders accurate read mappings to the reference genome, especially if different strains or closely related species are used. That, together with sequencing errors which are common even in SOLiD reads (despite the double interrogation) makes it hard to distinguish a true sequence polymorphism from a missequencing event or a false alignment.

SHRiMP algorithms offer an interesting solution, because they allow for mapping of more dissimilar sequences and deciding on the value of the match based on alignment quality statistics. We were able to test different stringency filtering conditions and choose for the most balanced approach (with 'normodds'≥ 0.7 cutoff value) in terms of gene calling and expression level, limiting false positives. Overall we reported 8,029 and 7,777 expressed genes for male and female gametes respectively. Based on this filtered, high quality alignment count data we performed a differential expression analysis using the edgeR package for R. Proper normalization is a crucial step in recovering relevant information about gene expression. In this study we used the TMM method (implemented in edgeR) that adjusts for the composition of RNA-Seq data rather than gene length. As a result we found 4117 differentially expressed genes between male and female gametes, which we then used for a functional study.

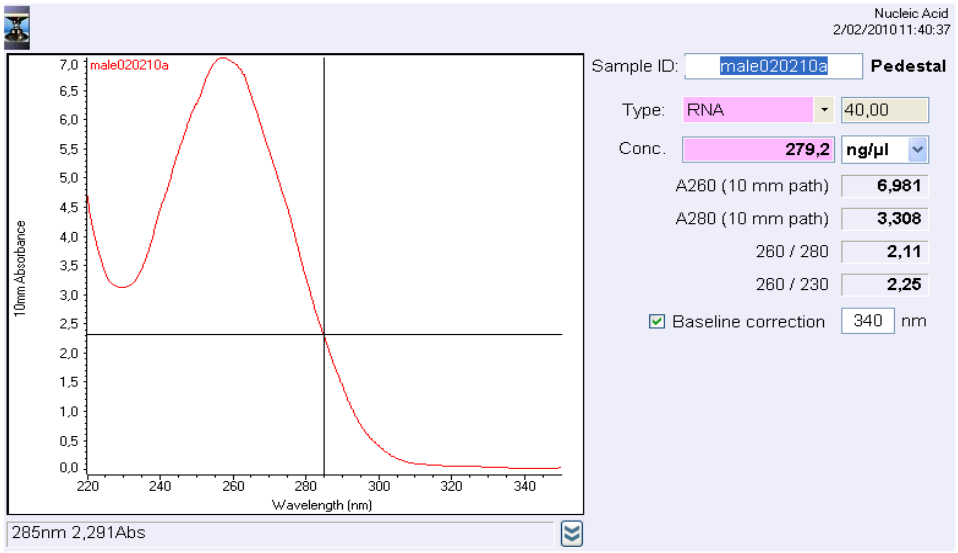
Overall the performed analysis and quality checks prove the reliability of the data set and make it a good basis for functional analysis of transcriptomes in male and female gametes (described in Chapter 4).

Supplemental data:

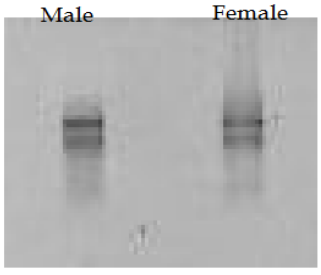
a)



b)



c)



Supplemental Figure 1. Quality of RNA used for transcriptome sequencing.

NanoDrop spectral data for female (a) and male (b) gamete total RNA. c) RNA integrity check (200ng) on a standard 1% DNA agarose gel.

Chapter 4

Uncovering the genetic basis for early isogamete differentiation: a case study of Ectocarpus siliculosus

Lipinska A.[1)]; D'hondt S.[1)]; Van Damme E.J.M.[2)]; De Clerck O.[1)];

[1)] Phycology Research Group and Center for Molecular Phylogenetics and Evolution, Ghent University, Krijgslaan 281, Building S8, 9000 Ghent, Belgium
[2)] Department of Molecular Biotechnology, Laboratory of Biochemistry and Glycobiology, Ghent University, Coupure Links 653, 9000 Gent, Belgium

Modified from submitted article*: BMC Genomics

* A.L. contributed to this paper by preparing the NGS libraries, performing the analyses of RNA sequencing data and drafting the manuscript.

Abstract

Background

The phenomenon of sexual reproduction characterizes nearly all eukaryotes, with anisogamy being the most prevalent form of gamete discrimination. Since dimorphic gametes most likely descend from equal-sized specialized germ cells, identifying the genetic bases of the early functional diversification in isogametes can provide a better understanding of the evolution of sexual dimorphism. However, despite the potential importance to the evolutionary biology field, no comprehensive survey of the transcriptome profiling in isomorphic gametes has been reported hitherto.

Results

Gamete differentiation on the genomic level was investigated using *Ectocarpus siliculosus,* a model organism for brown algal lineage which displays an isogamous sexual reproduction cycle. Transcriptome libraries of male and female gametes were generated using Next Generation Sequencing technology (SOLiD) and analyzed to identify differentially regulated genes and pathways with potential roles in fertilization and gamete specialization. Gamete transcriptomes showed a high level of complexity with a large portion of gender specific gene expression. Our results indicate that over 4,000 of expressed genes are regulated between male and female, including sequences related to cell movement, carbohydrate and lipid metabolism, signaling, transport and RNA processing.

Conclusions

This first comprehensive transcriptomic study of protist isogametes describes considerable adaptation to distinct sexual roles, where functional anisogamy precedes morphological differentiation. Several sex-biased genes and pathways with a putative role in reproduction were identified, providing the basis for more detailed investigations of the mechanisms underlying evolution of mating types and sexual dimorphism.

Keywords:

gamete, transcriptome, *Ectocarpus*, sexual reproduction, isogamy, brown alga, signaling

1 Introduction

Sexual reproduction encompasses the fusion of two specialized haploid cells to form a zygote. Phylogenetic analyses suggest that sexual reproduction arose already in the common ancestor of all eukaryotes(Ramesh et al., 2005), implying the existence of selective forces that gave an advantage to sexuality. Despite the costs and major challenges presented to the reproductive cells, over 99.9% of eukaryotes engage in sex (Stearns, 1987), which inspired generations of biologists to study this widespread phenomenon from physiological, molecular and evolutionary perspectives.

The existence of two gamete types and the tendency for gamete dimorphism remains an intriguing puzzle as anisogamy characterizes nearly all plants and animals. Oogamy (large eggs and small sperm) seemed to be the course of evolution from isogamy (equally-sized gametes) and arose independently in many groups of eukaryotes; however numerous species continue to reproduce with isomorphic germ cells (Kirk, 2006). Although substantial research addressed gene expression profiles in reproductive cells of flowering plants and animals (Ohnishi et al., 2011), and studies of the mating locus in Volvocine algae shed light on the transition towards oogamy(Umen, 2011), not much is known about how the global patterns of sex-biased gene expression were shaped throughout the evolution of mating types and transition towards anisogamy. Such studies are important, because most of the evolutionary models accept the existence of two specialized mating types upon which the evolution of gamete size was superimposed (Hoekstra, 1987). Therefore, a detailed characterization of transcriptional adaptation in equally-sized gametes would bring a better understanding to the mechanisms underlying evolution of sexual dimorphism. In this respect, brown algae (Phaeophyceae) with their broad spectrum of gamete copulation forms are suitable subjects to test various hypotheses (Bell, 1997).

Brown algae are a large group of multicellular, photosynthetic organisms, which evolved 150-200 million years ago. Distant to land plants and animals, they developed complex multicellularity independently from other major clades(Cock et al., 2010). This polymorphic group hosts seaweeds of a vast range of sizes, ecological niches and with an unmatched diversity of life cycles and fertilization strategies ranging from isogamy over anisogamy to oogamy (Bell, 1997) (Silberfeld et al., 2010). Despite the evolutionary distance, brown algae share many common features with land plants, which first brought much attention to eggs and zygotes of brown macroalgae, due to their large size and abundance, as a material to study the regulation of early development in plants (Brownlee et al., 2001). Other studies on reproduction have focused mainly on networks of signals that are associated with gamete attraction, recognition and fertilization success(Brawley, 1992). Although a large amount of

research concerns brown algal biology, many aspects remain poorly explored, providing excellent opportunities for new discoveries.

In recent years, following the selection of *Ectocarpus* as a model for the brown algae, a considerable effort was invested in the development of genomic and genetic tools for this organism, among which was the assembly and analysis of the complete genome sequence (Cock et al., 2012). *Ectocarpus* is a small filamentous alga, characterized by a haploid-diploid life cycle with isogamous sexual reproduction where flagellated gametes are still morphologically, but no longer physiologically, identical. Female gametes are distinguished by a short swimming period preceding settlement, flagella digestion and pheromone release (Müller, 1967). Fertilization takes place immediately after recognition by gender specific sex-receptors present on the egg surface and the male anterior flagellum. However, the dynamics and regulation of the mechanism driving male and female gamete differentiation and adaptation to fulfill their specific functions remain largely unexplored.

Here we describe the transcriptional networks specific to the gametes of both sexes. Using AB SOLiD 3 Next Generation Sequencing technology we generated whole RNA profiles of reproductive cells of E. siliculosus and determined the gender-specific regulation of the major metabolic pathways. The results present a first comparative gamete transcriptome analysis of any protist and provide an overview of the genes that contribute to the gametes' cellular identity and function.

2 Materials and methods

Culture conditions and gamete harvesting

Ectocarpus siliculosus (Ectocarpales, Phaeophyceae) unialgal strain NZKU 1-3 male gametophyte and NZKU 32-22-21 female gametophyte (origin Kaikoura, New Zealand) were cultivated at 12^0C in natural sea water enriched with modified Provasoli ES (West and McBride, 1999) with 14h light/10h darkness cycles (30 µmol × m^{-2} × s^{-1} flux density). To induce gamete release fertile gametophytes were transferred to Petri dishes with residual water only and kept overnight at 4^0C in the dark. Gamete release was induced by immersing cultures in PES in direct light at room temperature. Gametes were collected using a micropipette, transferred into 1.5ml Eppendorf tubes and pelleted at 5,000×g for 5 minutes. Gamete pellets were flash-frozen in liquid nitrogen and stored at -80^0C before RNA extraction.

RNA extraction and sequencing

Total RNA was isolated using an XS RNA extraction kit (Machery-Nagel) or RNeasy Plant Mini kit (Qiagen) according to manufacturer's instructions. An additional DNase digestion step

was performed in solution with RNase Free Turbo DNase (Ambion). The concentration and purity of all samples was measured with a Nano-Drop spectrophotometer (ND-1000, Thermo) and the sample integrity was checked on a 1% agarose gel. Approximately 20µg of total RNA from each type of gamete was rRNA depleted and shredded prior to cDNA synthesis using the SOLiD™ Total RNA-Seq Kit. Male and female samples were barcoded and prepared cDNA libraries were pooled and sequenced with a SOLiD 3 System (Applied Biosystems) at Cofactor Genomics (Missouri, USA). For more details on sample processing see Chapter 3.

Mapping to the reference genome and differential expression analysis

Reads were mapped to the reference genome (Sterck et al., 2012) using SHRiMP2 (Rumble et al., 2009). The statistical significance of top scoring hits was calculated using the Probcalc module of SHRiMP2 and only unique mappings with 'normodds' value >=0.7 were selected. Additionally Tophat software (Trapnell et al., 2009) was used to identify reads mapped in exon-exon splice junctions. Only exon mapped reads were considered in further analysis. Read processing involved filtering based on the number of reads per CDS, the covered length, and those with less than 5 reads mapped or covering less than 51 bp were discarded. These data were compiled into the gene expression table that served as input into the edgeR package for R (Robinson et al., 2010). Library normalization was done using the trimmed mean of M-values method (TMM) (Robinson and Oshlack, 2010) and Exact-Test was used to determine differentially expressed genes with P=<0.01 and FDR=<0.05 (see Chapter 3 for more ditails).

GO and KEGG enrichment analysis

To classify expressed genes, all sequences were annotated with KEGG orthology using KOBAS (Wu et al., 2006b) and Gene Ontology (GO) categories using Blast2GO (Conesa et al., 2005). These automatic annotations were used to investigate overrepresented pathways and GOs by comparison of individual libraries to all annotated genes in *Ectocarpus*. Over-expressed KEGG pathways were identified using the KOBAS web-platform (Wu et al., 2006b) and a hypergeometric test with Multiple Testing Correction of FDR (Benjamini and Hochberg, 1995). Over-represented GO terms were identified with Blast2GO and Fisher's Exact Test with Multiple Testing Correction of FDR (Benjamini and Hochberg, 1995).

Validation of RNA-Seq data by qRT-PCR

Quantitative real-time PCR was used to validate differential expression of ten selected genes (Table 1) and primers were designed using Primer3 software (Rozen and Skaletsky, 2000)

with default settings (Table 2). cDNA synthesis was carried out on 1µg of total RNA samples using oligo(dT)$_{12–18}$ primer (Invitrogen) and GOScript reverse transcriptase (Promega) according to the manufacturer's instructions. The qPCR reactions were performed in a 384-well thermocycler (LightCycler 480, Roche) with SYBR green chemistry (LightCycler 480 SYBR Green I Master mix, Roche) using listed conditions: 15 min at 95^{0}C, followed by 40 cycles of 15 sec at 95^{0}C, 20 sec at 50^{0}C, and 30 sec at 72^{0}C. Two biological replicates were run for both male and female cDNA samples and each sample was technically duplicated. Amplification specificity was measured with a melting curve by heating the sample from 65 to 97^{0}C and the product size was checked on a 1% agarose gel with GeneRuler™ 1 kb DNA Ladder (Fermentas). Absence of contaminating genomic DNA was checked with No-RT control PCR prior to cDNA synthesis. Normalization genes were selected using geNorm (Vandesompele et al., 2002) and the relative gene expression values were calculated in qBASE v.1.3.5. (Hellemans *et al.*, 2007).

Table 1

PCR primers used in this study for Real-Time PCR experiments

	Oligonucleotides		
Gene name	Forward	Reverse	PCR product length [bp]
Esi0102_0070	CTCAGCACTGCAGTCGTTAC	CGCGATCCAAGTGTACAAGG	166
Esi0067_0029	GCTGAAGTATCTCGACGGGA	TCTCATCGTACGGTCAACCC	220
Esi0069_0059	GAGATGCAACAACGTCGAGA	TCGAACGTGTTGTTGGTGAT	249
Esi0101_0018	AGATCAAGCTGGACAGGC	TGTGTATCGCAGTTCTCATT	253
Esi0104_0023	CCAACGCTCAGGTTCGCA	CCGTCCATGGCTCTCTCT	220
Esi0130_0068	ATCGGGGCCTTTCTCTCC	TGAAGGGAAGATCGCGATTC	147
Esi0418_0017	TTTGAGGGTGGCAAATAACC	CGTGTTTCTCTCCCTTCTCG	212
Esi0123_0020	CCTCCCTACGTCACCAAGAA	CACATCTTGTCGTCGTGCTT	239
Esi0161_0002	ACACAAGCCATTCCGATCAT	AGCGGGTACAACCATAAACG	182
Esi0098_0063	ATTGGCGTCGGGTTGTACT	TACCTTTCCGCATTGTGAGC	163
Esi0298_0008	ATGTCCGAAGACATGCAACA	TGGGTAACGTAGGACCCAAA	167
Esi0072_0068	GAACCACGGAAGGAACAAGA	GGAGGGCGTAGTTGTCGAAC	176

3 Results and Discussion

3.1 Next Generation Sequencing and mapping of the Ectocarpus transcriptome

Sequencing of rRNA-depleted total RNA of *Ectocarpus* gametes yielded more than 36 million 50bp reads for the male and 28 million 50bp reads for the female sample. At this sequencing depth we found at least five non-clonal reads uniquely aligned to 8,029 and 7,777 out of 16,239 annotated genes (male and female respectively). The overviews of the mapping procedure and quality evaluation are described in Chapter 3.

3.2 Gametes have unique transcriptional profiles

A Venn diagram (Fig. 1) displays common gene expression between gametes and representative EST libraries(Dittami et al., 2009) of vegetative gametophyte and sporophyte tissues (corresponding to 9,163 annotated genes). Approximately 70% of the EST sequences were shared by gametes and vegetative tissues. This is not surprising, since non-fertilized *Ectocarpus* gametes are capable of parthenogenesis and development into functional parthenosporophytes (Müller, 1967). However, almost one-third of the gamete-expressed transcripts were found exclusively in the reproductive cells. Gene ontology (GO) analysis of this subset indicated that sequences related to signal transduction, RNA modification and localization and microtubule based movement were significantly enriched ($p<0.01$). The high contribution of gamete-specific mRNAs within the whole transcriptome pool highlights the potential significance and regulatory specialization of this subset.

We also looked at the top 100 most expressed genes and manually grouped them into functional categories based on gene annotation information (Supporting information Table S2). The two largest clusters were composed of genes related to carbohydrate metabolism including cell wall biosynthesis (11 Female; 11 Male) and protein turnover (7 Female; 8 Male). However, the majority of the most abundant transcripts were of unknown function (60 Female; 55 Male).

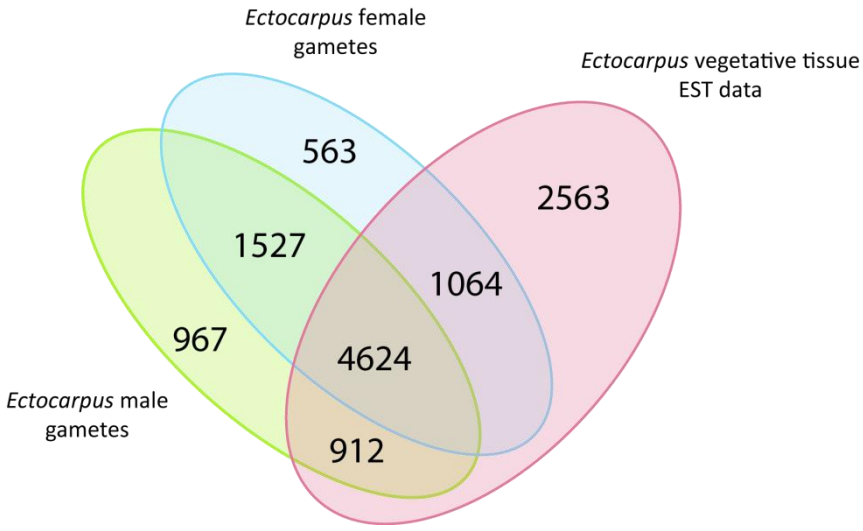


Figure 1. Comparison of gene expression in gametes and vegetative tissue. Venn diagram showing the overlapping genes that were found expressed in male gametes (8,029), female gametes (7,777) and *Ectocarpus* vegetative tissue EST represented genes (9,163) (Dittami et al., 2009).

3.3 Classification of gamete-expressed genes with automatic annotations

Around 62% of the gametes' transcripts could be assigned with a Gene Ontology category using Blast2GO (E-value < 1e-05). Statistical analysis marked biological processes including translation and gene expression, auxin biosynthesis, proteolysis, transport, localization and regulation of signal transduction as significantly overrepresented in both gamete types (FDR<10%). Additionally, we found that transcripts related to vesicle-mediated transport, lipid metabolism and iron/sulfur cluster assembly were significantly overrepresented in the female library, whereas sequences related to pigment biosynthesis were enriched in the male library. In the "cellular component" category, differences were observed in mitochondrion, nucleus, vesicle membrane and Golgi related components (overrepresented in female library) and chloroplast stroma (overrepresented in male library) (additionally see Fig. 2 for enrichment in Molecular Function and Supporting information Table S3 for full overview).

In general, genes related to photosynthesis were underrepresented in the transcriptome of both gametes. Female gametes were also deficient in sequences related to DNA metabolic processes (e.g., protein-DNA complex assembly, nucleosome organization) as well as microtubule-based movement and male gametes had underrepresented genes in the cellulose binding group.

Based on significant similarity (E-value < 1e-05) we also assigned 2,418 and 2,243 Kegg orthology terms to the gamete's expressed proteins (male and female respectively) using the KOBAS server (Wu et al., 2006a). A significant proportion of the transcripts in females (FDR<10%) were associated with ribosome, spliceosome and endocytosis. Ribosome was the only valid pathway overrepresented in male gametes with the given threshold (FDR<10%).

3.4 Differential gene expression analysis

The preferential expression of genes belonging to a specific functional category became more evident when differentially expressed genes were considered. Using the edgeR package for R (FDR of 5% and a fold change >= 2) 4,117 genes were identified as differentially expressed between male and female gametes (Chapter 3). Apparent enrichment could be seen in particular with the categories microtubule based movement, vesicle trafficking, ion dynamics, cell wall biosynthesis, transcription and translation regulation, and signaling related genes, which are described below (Fig.3 and 4, Supporting information Table S5; for details about involved genes see Supporting information Table S6).

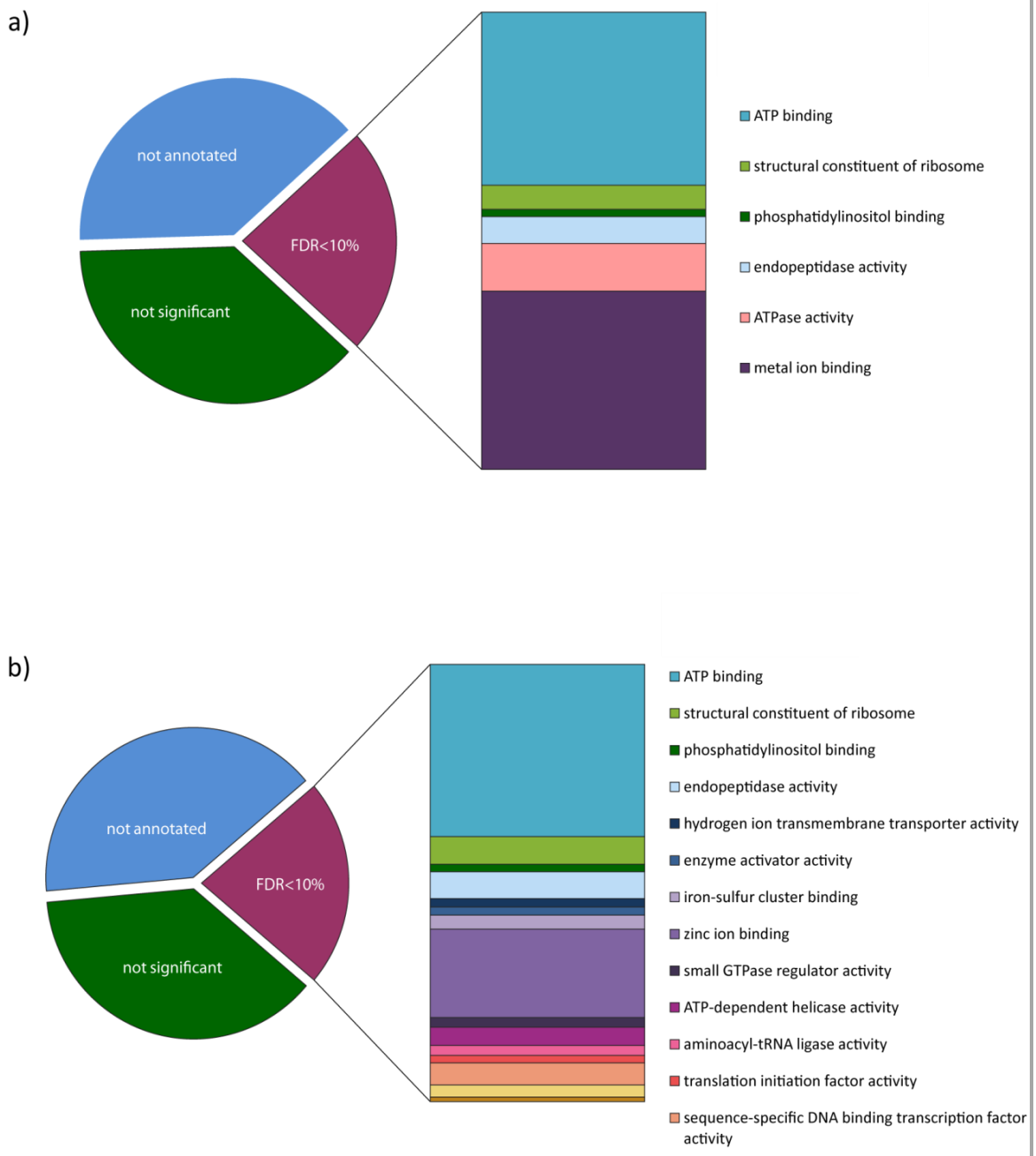


Figure 2. Significantly enriched (FDR<10%) Gene Ontology groups of gametes expressed genes according to molecular function. a) Male gametes expressed genes; b) Female gametes expressed genes.

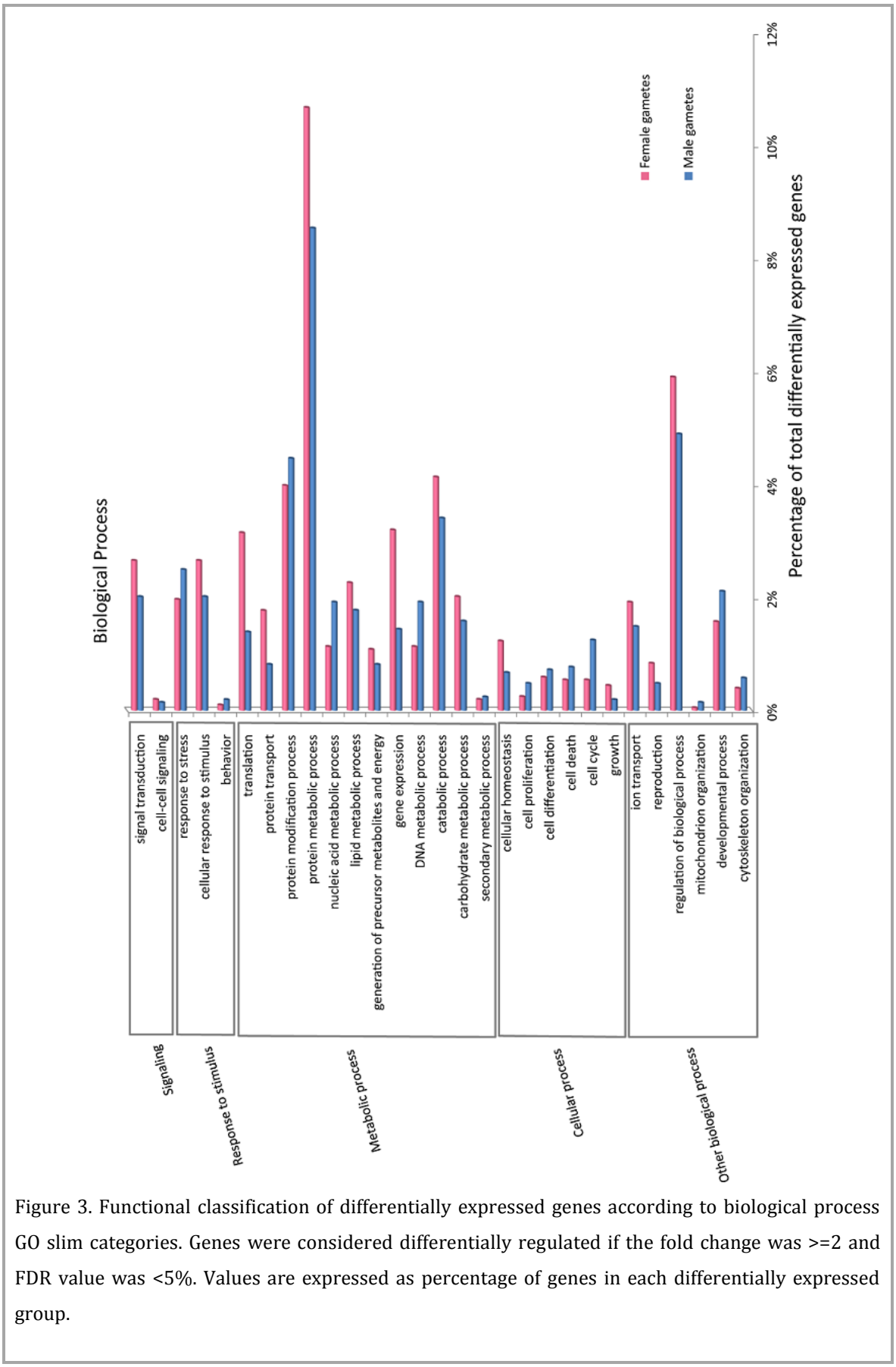


Figure 3. Functional classification of differentially expressed genes according to biological process GO slim categories. Genes were considered differentially regulated if the fold change was >=2 and FDR value was <5%. Values are expressed as percentage of genes in each differentially expressed group.

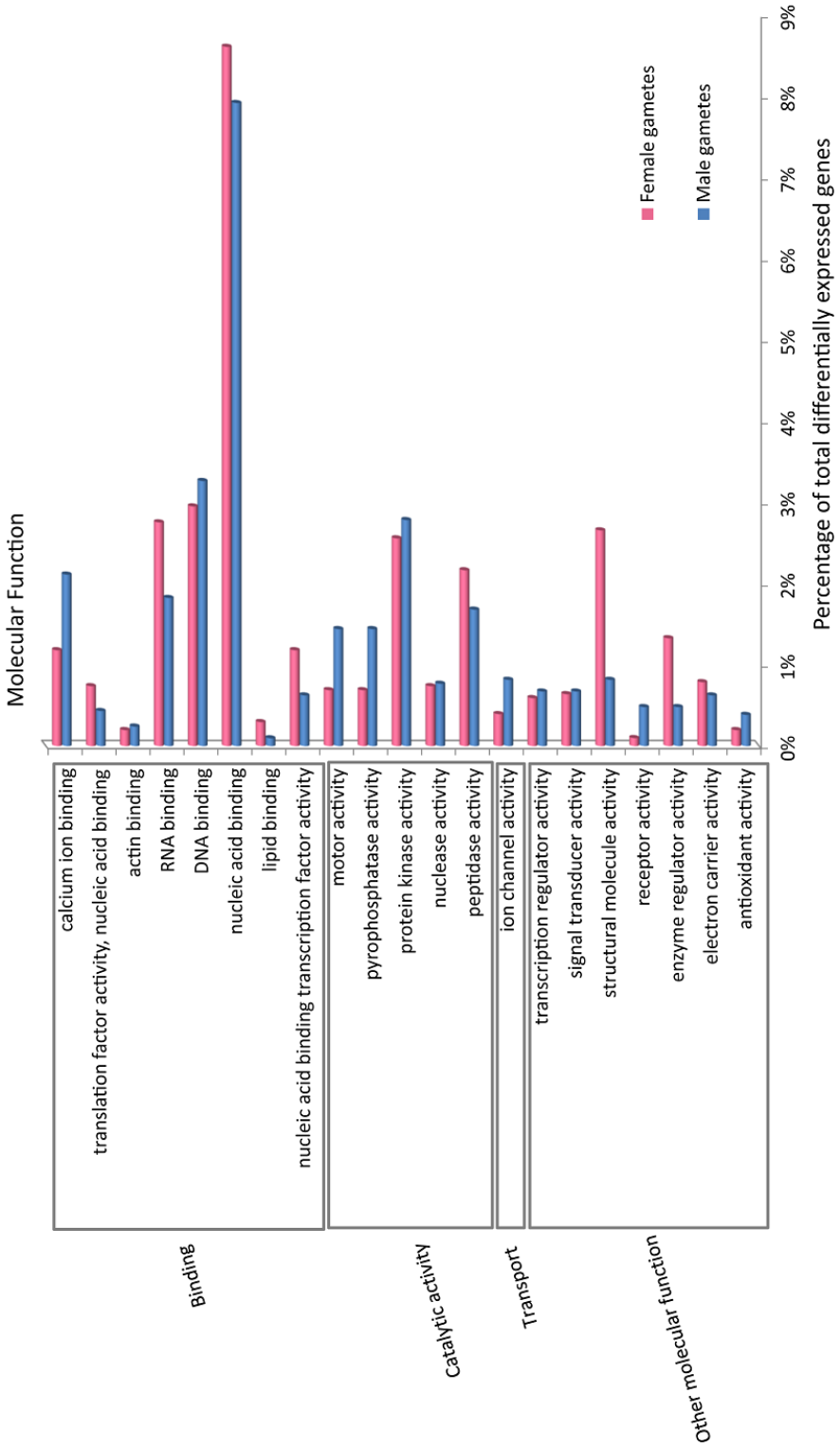


Figure 4. Functional classification of differentially expressed genes according to molecular function GO slim categories. Genes were considered differentially regulated if the fold change was >=2 and FDR value was <5%. Values are expressed as percentage of genes in each differentially expressed group.

Microtubule based movement

The terms overrepresented in male gametes are assigned primarily to microtubule based movement (GO:0007018), which can be associated with the sperm active swimming behavior. Besides the canonical role in locomotion, *Ectocarpus* flagella are also important sensory organs involved in chemosensation and gamete recognition (Schmid, 1993). Several genes belonging to the intraflagellar transport (IFT) and motor protein families were overexpressed in male gametes. IFT proteins are macromolecular rafts responsible for the assembly and maintenance of the flagella (Pedersen and Rosenbaum, 2008) and the deposition of mastigonemes on the flagella surface after their assembly in the ER (Bouck, 1971). Growing evidence suggests that IFT plays a more direct role in cilia-mediated signaling(Wang et al., 2006). Among other genes linked to the gamete flagella, we found members of the Sexualy Induced Gene family – Sig1 and Sig2-like gene, with Sig1 among the highly overexpressed genes in male gametes. This family of proteins was first described during the onset of sexual reproduction in the diatom *Thalassiosira* (Armbrust, 1999) and later also in other Stramenopiles (Honda et al., 2007). Sig proteins are located to the mastigonemes (Yamagishi et al., 2009), but their function remains unresolved. However, striking diversification of Sig1 between closely related species of *Thalassiosira* (Armbrust and Galindo, 2001) and some evidence of positive selection acting on this gene (Sorhannus and Kosakovsky Pond, 2006), may suggest a role in gamete recognition.

Ion dynamics

Potassium channel activity (GO:0005267) and calcium-activated potassium channel activity (GO:0015269) were significantly overrepresented in male gametes. These might be related to the sperm chemotaxis in analogy to the gene network triggering motility response to stimuli in sea urchin sperm (see(Darszon et al., 2008)(Neill and Vacquier, 2004) for a review). It has been shown that cyclic nucleotide messengers and changes in K^+ ion dynamics lead to hyperpolarisation of the cell membrane and activation of the Na^+ and Ca^{2+} influx in sperm(Granados-Gonzalez et al., 2005)(Strünker et al., 2006). Recently, a transient increase in Ca^{2+} in the flagellum was directly visualized during chemotactic orientation in ascidian sperm using a fast Ca^{2+} imaging system (Shiba et al., 2008). Previous studies on *Ectocarpus* pheromone response confirm participation of free Ca^{2+} in sperm navigation, since concentrations below 10^{-7} M caused male gamete immobilization despite the presence of the attractant (Maier and Calenberg, 1994). Highly upregulated male genes, with a homology to the *Strongylocentrotus* sperm pathway, included three cyclic nucleotide binding K^+ channels with similarity to the TetraKCNG channel (BLAST E value >1e-20), a Na+/H+ exchanger, an adenylate cyclase, a sperm hyperpolarization-activated and cyclic nucleotide-gated channel

(BLAST E value >1e-15) and a similar voltage gated-calcium channel (BLAST E value >1e-33).

Cell wall/polysaccharide biosynthesis

Members of different carbohydrate biosynthesis pathways were predominantly upregulated in female gametes, which relates to primary cell wall biogenesis minutes after fertilization (Callow et al., 1978a). All enzymes involved in alginate synthesis (Michel et al., 2010) and three *Ectocarpus* cellulose synthases (CESAs) are highly overexpressed in female gametes, which is apparent with the brown algal cell walls being composed of alginate with a minor fraction of cellulose (Kloareg and Quatrano, 1988). Apart from alginate and cellulose, sulfated fucans and phenolic biomolecules (phlorotannins) are secreted into the expanding cell wall (Callow et al., 1978b). We can assume that these compounds are also synthesized in gametes, since all fucosyltransferases (except from family GT24) and sulfotransferases (STs) (except Clade B) (Michel et al., 2010) were transcribed with some gender specificity. It is worth noting that one sulfotransferase (Esi0032_0064), related to metazoan STs involved in the biosynthesis of glycosphingolipids, was characterized by a much higher expression comparing to other STs in both gametes. Associated with lipid rafts, glycosphingolipids may act as intermediates in signaling the flow from outside to inside the cell (Schnaar et al., 2009). Sulfated fucans and galactans are also reported to be involved in sea urchin fertilization. They act as inducers of the sperm acrosome reaction (Alves et al., 1997)(Mourão, 2007) by binding to the sperm receptor REJ (Receptor for Egg Jelly), a homolog of the human polycystin protein (Gunaratne et al., 2007). Interestingly, five of the expressed sulfotransferases were specific only to female gametes and five polycystin/REJ-like proteins (IPR002859) were expressed exclusively in males.

Vesicle trafficking

Gametes and spores of *Ectocarpus* can be characterized by the presence of several active Golgi bodies (Baker and Evans, 1973a)(Maier, 1997). Similar observations were made with early electron scanning photographs of *Fucus* unfertilized eggs, which show a characteristic, rough surface due to protrusion of the cytoplasmic vesicles beneath the plasma membrane (Callow et al., 1978a). These findings are reflected in the upregulated genes of female gametes, since clathrin coat proteins (GO:0030118) constituting vesicles travelling from the Golgi apparatus to the plasma membrane (Kirchhausen, 2000) are highly abundant. The same is observed with the retromer complexes (GO:0030904), which assemble on early endosomes and are involved in transport back to the Golgi apparatus. Additionally, the Rab protein signal transduction pathway (GO:0032483), including the Arf family, which are coat-recruitment GTPases (for a recent review, see (Novick and Zerial, 1997)(Zerial and McBride, 2001)) and

dynamins, which are necessary for pinching the vesicles (Kirchhausen, 2000), are upregulated in female gametes.

The Golgi complex is the major site for polysaccharide synthesis including alginates, sulfated fucans and phlorotannins of the algal cell wall, which are transported in vesicles to the plasma membrane (Callow et al., 1978b)(Schoenwaelder and Wiencke, 2000). Thus, the upregulated pathways of cellulose biosynthesis and vesicular transport can support primary cell wall biogenesis. It is also possible that this secretory activity may be important for biosynthesis of the adhesive substance required for gamete attachment to a substrate (Charrier et al., 2008).

Translation and transcription regulation

Gene expression in gametes is developmentally regulated and stage specific, and thus requires a precise and well-coordinated program of transcription regulation. Overall *Ectocarpus* possesses 14 families of transcription factors (TFs): HS; Myb; bZIP; Zn_finger C2H2-type, CCCH-type, TAZ-type; bHLH; fungal TRF; CCAAT-binding; AP2-EREBP; TAF9; E2F-DP; CXC/tesmin; Sigma-70r2/r3/r4 (Rayko et al., 2010). Gametes express TFs across most of the described families. In particular, we find members of Heat shock (HS) factors, fungal TR and CCAAT-binding overrepresented in females and most MYB genes overexpressed in males. Interestingly several MYB factors were also relatively high expressed in *Arabidopsis* sperm cells (Borges et al., 2008). Among the highest expressed TFs we find two MYB factors (Esi0038_0132, Esi0212_0014), Zinc-finger C2H2-type factor (Esi0226_0040) and two fungal TRFs (Esi0008_0230, Esi0348_0008). Esi0212_0014 shares 53% identity (8e-37) with *Arabidopsis* MYB98 which controls the formation of specific features within the synergid cell during female gametophyte development (Kasahara et al., 2005)(Punwani et al., 2007).

Another interesting transcription factor family described by (Cock et al., 2010) is the NIN-like protein, coded by nine genes in *Ectocarpus*. NIN-proteins are required for symbiosis between legumes and nitrate assimilating bacteria, and a subfamily of NIN, the minus dominance proteins (MID), are expressed during gametogenesis in volvocine algae and determine the minus mating type (Ferris and Goodenough, 1997). It was suggested, that the NIN-like family might have a role in the life cycle or in the mating type determination in *Ectocarpus* gametes(Cock et al., 2010). Indeed eight members of this family are found in our data, with two of them being specific for female gametes. It is also worth noting, that the most highly expressed NIN factor in both gametes (Esi0013_0140) was significantly downregulated in the *immediate upright* mutant, which is defective in sporophyte development (Peters *et al.*, 2008).

Protein metabolic processes, in particular biosynthetic pathways (ribosome and translation

related) are significantly enriched ($p<0.01$) in female DE genes. Additionally, *Ectocarpus* was shown to have a micro RNA post-transcriptional regulation system, where most of the mature miRNA bare a signature preferred by the plant Argonaute-1 protein (AGO1) (Cock et al., 2010). The genome contains one AGO1 protein which is medium expressed in both types of gametes. The Argonaute-miRNA are known to silence transcription, trigger target destruction, or inhibit translation, and growing evidence supports their role in germline development (Khurana and Theurkauf, 2010)(Wuest et al., 2010). Moreover, several potential target sequences including members of the ROCO family GTPases and other proteins containing leucine-rich repeat (LRR) domains (Cock et al., 2010) are expressed in gametes. Since gametes are vulnerable targets for pathogen attack (Dixon et al., 2000)(Müller et al., 2000), these proteins might be involved in algal immune response to disease (Zambounis et al., 2012).

Signaling pathways

Gamete transcripts are enriched in Ras GTPase superfamily genes (Ran, Ras, Rab, Rho and Arf). These signaling molecules are binary switches in crucial cellular processes including growth, differentiation and survival (Rojas et al., 2012). Rab and Arf are particularly important in membrane trafficking and are enriched in female gametes (see Vesicular transport). The Rho family is involved in signaling networks that regulate actin, cell cycle progression, and gene expression. Noteworthy RAC, a Rho family GTPase, and its positive effector RhoGEF are upregulated in females, whereas RhoGAP, a negative regulator, is highly expressed in male gametes. Rho genes, RhoGDI (guanine nucleotide dissociation inhibitor of Rho) and a RhoGAP were also expressed in *Arabidopsis* sperm, where no RhoGEFs were found (Borges et al., 2008). The precise function of Rho signaling in sperm and egg remains to be revealed, but substantial work has been done by Kumakiri *et al.* (Kumakiri et al., 2003), showing a role in initial sperm-egg fusion in mouse. In the study, *Clostridium difficile* toxin B inhibited sperm incorporation probably by disturbing actin filament reorganization regulated by Rho GTPases. Rac1 seemed to be strongly expressed in mouse eggs and located in the cortical ooplasm. The process of sperm–egg fusion would be initiated immediately after sperm binding by membrane receptors that in turn would activate Rho proteins by RhoGEFs. Membrane receptors activating RhoGEFs include G protein-coupled receptors, such as the lysophosphatidic acid (LPA) receptor, the growth factor receptors with a tyrosine kinase domain, such as EGF receptors, and surface proteins such as integrins (Schmitz et al., 2000). Members of all these receptor families could be identified in gametes, but the specific recognition protein involved remains unknown.

Other regulators of Rho GTPase activity (GO:0032319) are expressed in male gametes,

including two Target Of Rapamycin (TOR) kinases (TOR1 and TOR2). TOR is a nutrient-sensitive, central controller of cell growth and aging, which was linked to the actin cytoskeleton via a signaling pathway containing a Rho GTPase (Schmidt et al., 1997). Raptor and FKBP12, TOR associated proteins, are also expressed, but no RAG GTPases that promote intracellular localization of TOR were present. *Ectocarpus* has two TOR kinases, which in yeast act in two different pathways (for a recent review see (Loewith and Hall, 2011)). One pathway involved in cell growth in response to nutrient availability is shared between TOR1 and TOR2; however, TOR2 has additional, unique functions in sphingolipid synthesis, endocytosis and polarized organization of the actin cytoskeleton. Fold change expression analysis shows TOR2 to be upregulated in male gametes (FC=8) compared to TOR1 (FC=1.2), which could point to the importance of the second branch of TOR2-signaling in males, especially with high expression of glycosphingolipid-related ST (see Polysaccharide biosynthesis).

Ras GTPases influence transcription of genes involved in cell growth and division by activating protein kinases, such as the mitogen-activated protein (MAP) kinase. Several members of the family were detected in *Arabidopsis* sperm cells and some are sperm specific, implicating the existence of unique signaling pathways (Borges et al., 2008). Out of five MAPKs expressed in gametes, two were upregulated in males. We found also one MAPK related serine/threonine protein kinase specific to male gametes, with homology to the LF4 gene (MAPK) localized in *Chlamydomonas* flagella (Berman et al., 2003). This protein is involved in a signal transduction cascade controlling flagellar length.

Another effector activated by Ras is phosphoinositide-3-kinase (PI3K). There are two putative PI3Ks in *Ectocarpus*, one highly expressed in both types of gametes and one upregulated in males. The product of PI3K, phosphatidylinositol 3-phosphate, plays an important role in regulating membrane trafficking. Additionally, we identified two enzymes necessary for phosphoinositide-mediated signaling which were enriched in male gametes, phosphatidylinositol 4-kinase and 1-phosphatidylinositol-4-phosphate 5-kinase, involved in synthesis of phosphatidylinositol 4,5-bisphosphate (PIP2). PIP2 is a minor constituent of the plasma membrane, where it functions as an intermediate in a number of signaling pathways, including G protein-coupled receptor (GPCR) signaling. The sperm-induced breakdown of the PIP2 via activation of phospholipase C is considered to be the major reaction of fertilization (Sillers and Forer, 1985)(Halet et al., 2002). The importance of the PIP2 secondary messenger system in sexual reproduction was shown in echinoderms eggs, where it regulates Ca^{2+} release at fertilization and controls the slow polyspermy block (Lee and Shen, 1998)(Townley et al., 2006). PIP2 was also abundant in the plasma membrane and the flagellar membrane of *Chlamydomonas eugametos* gametes, indicating involvement of the phosphatidylinositol-

calcium signaling system during mating, which could be activated by binding of cell-cell recognition receptors (Brederoo et al., 1991). Additionally, alternative Ca^{2+} gates like ryanodine receptors may be involved following the propagation of a calcium wave (Lee and Shen, 1998)(McDougall et al., 2000). One member of inositol triphosphate/ryanodine-type receptors is represented in *Ectocarpus* and found highly expressed in both types of gametes. Existence of inositol 1,4,5-trisphosphate-induced Ca^{2+} waves has been reported in *Fucus* embryos, emphasizing the importance of calcium signaling in response to a physiological stimulus (Goddard et al., 2000)(Coelho et al., 2002). One of the current hypotheses about sperm induced oocyte activation assumes stimulation of a membrane receptor that involves G protein signaling (Wilding and Dale, 1997). G-protein coupled receptors (GPCRs) are transmembrane proteins that utilize interactions with heterotrimeric G proteins (Gα, Gβ and Gγ) for downstream signaling and the pathway depends on the isoform of the α-subunit to which the receptor is coupled (Wettschureck and Offermanns, 2005)(Neves et al., 2002). Six paralogs of the Gα subunits (GPA) are found in the *Ectocarpus* genome and are all expressed in gametes. GPA4 and GPA6 are among the highly transcribed genes whereas the GPA3 and GPA4 are upregulated in male gametes. Moreover, three putative GPCR receptors are specific to male gametes and three partial GPCRs are upregulated in females. Substantial evidence supports a role of GPCRs in egg-sperm interactions during fertilization. For example, a G-protein coupled receptor located on the spermatozoa plasma membrane activates a signaling pathway responsible for the zona pellucida induced acrosomal exocytosis (Ward et al., 1992). A G-protein coupled cAMP transduction pathway is also involved in chemotaxis in human sperm (Spehr et al., 2004) and Gα proteins together with adenylyl cyclase were shown to be enriched in sea urchin sperm (Ohta et al., 2000).

The GPCR receptor family is a host to many pheromone receptors (Bockaert and Pin, 1999). It was shown that *Ectocarpus* sperm chemotaxis is stimulated in a similar manner as for pheromones by trifluoperazine (TFP) (Maier and Calenberg, 1994), which is an antagonist of dopamine/adrenergic G-protein coupled receptors. Thus, it might be possible that TFP activates the ectocarpene receptor which could belong to GPCR family.

Male and female gametes express genes related to Hedgehog and Notch signaling pathways involved in animal development (for a review see (Baron, 2003)(Ehebauer et al., 2006)). Although these pathways do not exist in a canonical form in non-metazoans, it has been shown that components such as the γ-secretase complex, Notchless and Hog/Hint domain proteins are of ancient origin (Bürglin, 2008)(Gazave et al., 2009) and new receptors seem to evolve by shuffling of pre-existing domains. The presence of Notch receptor building blocks is revealed in the *Ectocarpus* genome, however no homologues of the Notch receptor or its ligands *sensu stricto* have been found. A KEGG orthology analysis of gamete transcripts

identified Deltex, a Notchless homolog, Presenilin and Nicastrin from the γ-secretase complex, two histone deacetylase co-repressors and three co-activators of the DNA binding protein with one highly expressed putative histone acetyltransferase. Additionally, 16 genes with a Notch domain (IPR000800) including Esi0061_0098, described by Le Bail et al. (2011) as downregulated in the *Ectocarpus* developmental mutant – *etoile,* were present. Regarding the Hedgehog pathway, nine genes with similarity to Patched receptor (Ptc) of Hedgehog (containing both Patched (IPR003392) and SSD (IPR00731) domains) were abundant in gametes and one gene with Hint (Hedgehog/Intein N-terminal domain (SMART00306)) was low expressed only in females. Ptc and hint-domain proteins as well as Nicastrin and Notch-domain containing proteins were present during sexual reproduction in pennate diatom *Seminavis robusta* (Gillard, 2009), although their involvement in cell-cell interaction is unknown and awaits further research.

Pheromone biosynthesis

Brown algal pheromones are C-11 hydrocarbon compounds derived from fatty acids (Pohnert and Boland, 2002). Female gametes of *Ectocarpus* use arachidonic acid as a precursor of ectocarpene (Stratmann et al., 1993) and accumulate large reserves of phosphoglyceride PX, rich in arachidonic and eicosapentaenoic acid, in their plasma membrane (Schmid, 1994). The hypothesized synthesis pathway involves lipoxygenase to form a peroxidised lipid and hydroperoxide lyase (LOX) for cleavage at the peroxidized site to C-11 hydrocarbon and conjugated oxoacid as a by-product (Pohnert and Boland, 2002). GO and KEGG analyses revealed few lipoxygenases upregulated in female gametes, but no homologues of hydroperoxide lyases (HPL). Two Allene Oxide Synthases (AOS) indicated as putative hydroperoxide-lyases in pheromone pathway by Cock *et al*. (Cock et al., 2010) were also not expressed. Nevertheless, it could be possible that *Ectocarpus* LOX exhibits a double activity, like PpLOX1 from the moss *Physcomitrella patens*. Lipooxygenase form *Physcomitrella* combines the function of hydroperoxidase liase and acts on 18-22 carbon chains substrates (Senger et al., 2005). It is also significantly similar to Esi0424_0006 LOX from *Ectocarpus* (E=6e-29 (43%)).

Ectocarpus eggs are significantly enriched in genes related to glutathione (GSH), namely glutathione synthases and glutathione S-transferase. Glutathione as a radical scavenger prevents damage of cellular components caused by reactive oxygen species (ROS), such as peroxides (Sheehan et al., 2001). Thus glutathione synthesis might be female's cytoprotective strategy against oxidative damage in the presence of peroxidized lipids. A similar hypothesis was tested by Arab et al. (2006) where the influence of several fatty acids on the redox status and lipid peroxidation in human fibroblasts was investigated. Noteworthy, among all tested

lipids particularly arachidonic acid and conjugated linoleic acid induced glutathione synthesis to protect the cells oxidative balance. It is compelling in relation to the fact that arachidonic acid is accumulated in female gametes as a pheromone precursor. Additional incubation of a crude extract of the brown alga *Laminaria angustata* with arachidonic acid in the presence of glutathione slightly enhanced the reaction efficiency, probably through protecting the enzymes from oxidative inactivation (Boonprab et al., 2003). However, the endogenous level of GSH in *Laminaria* was suggested to be low.

Table 2

Validation of SOLiD based gene expression profiles by Real-Time PCR

Gene name	Gene function	Log_2 (Average relative expression)[a] SOLiD	RT-PCR	Validation (yes/no)
Esi0102_0070	Arf1, ARF family GTPase	-1.93	-2.05	Y
Esi0067_0029	long chain acyl-coA synthetase	-4.19	0.49	N
Esi0069_0059	Mannuronan C-5-epimerase	-2.98	0.06	N
Esi0101_0018	Tubular mastigoneme-related protein	6.16	4.72	Y
Esi0104_0023	GPCR-like protein	4.72	2.22	Y
Esi0130_0068	PKD/REJ-like protein	0.62	2.22	Y
Esi0418_0017	MORN motif precursor	4.95	32.19	Y
Esi0123_0020	hypothetical protein	-7.15	-31.61	Y
Esi0161_0002	Metal ion transporter-like protein	3.50	-32.10	N
Esi0098_0063	hypothetical protein	-4.92	-31.61	Y

[a] Relative expression is calculated as a ratio of expression levels in male/female gametes to indicate genes up- or downregulated in male gametes. A gene is considered differentially expressed if its relative expression is twofold or greater.

3.5 RT-PCR validation

Ten genes that were identified with a high level of significance were selected to confirm the RNA-Seq results via qRT-PCR. To find the best normalization genes for gamete libraries, we investigated the expression of housekeeping genes reported by Le Bail (Le Bail et al., 2008) for microarray experiments. After analysis with geNorm (Vandesompele et al., 2002) dynein and ribosomal protein 26S showed the smallest relative stability M-value (M=0.27) across male and female gamete samples and were selected for normalization. Real-time PCR results were in general consistent with the direction of relative expression changes obtained by RNA-Seq, with a Pearson coefficient R of log_2(Fold Change) equal to 0.53, indicating a positive correlation between qPCR and RNA-Seq data. However, differences in the exact fold change values were observed (Table 2). Furthermore, the most stable 'housekeeping genes' as identified by qPCR in a previous report (Le Bail et al., 2008) including ubiquitin conjugating enzyme (UBCE), alpha tubulin (TUA), actin related protein (ARP2.1) and translation

elongation factor 1 alpha (EFIa), showed only statistically non-significant relative changes of <1.5-fold (log2-ratio <0.58) in expression.

4 Conclusions

Here we provide the first to our knowledge, comparative analysis of protist gametes' transcriptomes. One of the key findings of this study is that *Ectocarpus* gametes equal the intricate transcriptomes of oogamous species (Engel et al., 2003)(Wuest et al., 2010). Most of the transcribed genes may not have an evident role before fertilization, nor are they necessarily translated, but they may be crucial during post-fertilization development as in plant and animal systems (Evsikov et al., 2006)(Xin et al., 2011). A large set of the expressed genes is common to somatic tissues, which implies their core metabolic functions and presumably also a role in the parthenosporophyte development. However, 4,117 genes in the *Ectocarpus* gametes' transcriptome are differentially regulated and one-third of the identified transcripts seem to be gamete specific, with primary functions in signal transduction and RNA processing. It is remarkable that within the morphologically identical isogametes, the transcriptome profile is substantially divergent, reflecting the early establishment of distinct sexual roles. Both males and females are able to regulate levels of mRNA engaged in many cellular processes. The female transcriptome is depleted in genes related to chromatin organization and enriched in genes with function in cell wall biogenesis, vesicular transport, lipid metabolism with pheromone synthesis, gene expression and signaling. In male gametes a significant part of the upregulated genes relates to microtubule based movement and ion flux as well as signal transduction. These results can be linked to previously described gamete characteristics in *Ectocarpus* and sister species, like active swimming in males with a tight relation to ion dynamics (Maier and Calenberg, 1994), vesicle protrusions in settled females (Callow et al., 1978a), pheromone production (Müller and Schmid, 1988) and DNA dispersion in egg nuclei (Callow et al., 1978a). In addition, our data confirm that transcripts related to cell wall biogenesis are deposited in female gametes before fertilization. We also revise the proposed pheromone pathway and imply the potential role of glutathione in maintaining the cell oxidative balance. The comparative RNA-Seq analysis presented here revealed a number of signaling pathways potentially involved in gamete recognition and fertilization. In particular, genes related to phosphatidylinositol signaling, GPCR receptors, REJ-like proteins and sulphonyltransferases were found, providing new insight into the mechanism of gamete coupling. Furthermore, the identification of differentially expressed transcription factors (like MYB or NIN-proteins) brings potential for discovery of sex specific gene expression regulators.

Taken together, we demonstrated a highly functional specialization in morphologically identical isogametes of *Ectocarpus*. Further insights into activated genes and pathways regulating gamete differentiation will result not only in better understanding of these reproductive cells and their interactions during fertilization but may also link sex determination to the formation of functional male and female gametes and shed light on the forces shaping the evolution of different sexes.

List of abbreviations

bp: base pair; PCR: polymerase chain reaction; TMM: trimmed mean of M values; GO: gene ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; FDR: false discovery rate; IFT: intraflagellar transport; CESA: cellulose synthase; ST: sulfotransferase; TF: transcription factor; TFP: trifluoperazine; GSH: glutathione; ROS: reactive oxygen species.

Additional data files

The following additional data are available with the online version of this paper:

Supporting Information Table 1. List of genes expressed in *Ectocarpus* gametes.

Supporting Information Table 2. *Ectocarpus* gametes' 100 most expressed genes.

Supporting Information Table 3. Gene Ontology enrichment of gametes' all expressed genes.

Supporting Information Table 4. List of gametes' differentially expressed genes.

Supporting Information Table 5. Gene Ontology enrichment of gametes' differentially expressed genes.

Supporting Information Table 6. Genes associated with upregulated pathways described in Results and Discussion.

Acknowledgements

This work was funded by BOF Grant 09/24J/117, Ghent University, Belgium. We gratefully acknowledge Dieter Müller for providing the *Ectocarpus* strain used in this study and for helpful instructions on culturing. We thank Erwan Corre for guidance and advice in setting up the bioinformatic analysis as well as Susana Coelho and Mark Cock for helpful discussions and words of encouragement.

Chapter 5

Molecular evolution of candidate male reproductive genes in Ectocarpus siliculosus

A.L. contributed to this chapter by performing DNA extractions, PCR experiments and sequence data analysis.

1 Introduction

It is generally established that natural selection shapes the observed genetic variation. However, the extent to which natural selection acts on genetic variants varies greatly between species (Bustamante *et al.*, 2005). Furthermore, natural selection can be problematic to quantify, mainly because the influence of variation on the individual's fitness is not easily measurable. Since mutations occur frequently and may accumulate in the population due to random factors, it is difficult to attribute the observed alleles either to genetic drift or diversifying selection. One way to distinguish between these two scenarios is to compare protein-coding DNA sequences between different species. The availability of large scale genomic data helps resolving this issue by identifying particular instances where genomic variation is a result of natural selection, at the same time helping to uncover important functional characteristics of investigated gene products.

The comparison of synonymous (d_S) and non-synonymous (d_N) substitution rates in protein-coding genes has proven to be a useful measure for molecular evolution. Selective pressure is determined by the non-synonymous/synonymous ratio ($\omega=d_N/d_S$). If non-synonymous mutations are deleterious, they would be present at lower rates resulting in $\omega<1$ (purifying selection). If the fitness is not influenced by synonymous or non-synonymous changes, the ω ratio will be close to 1 (neutral evolution). However, if non-synonymous mutations raise an individual's fitness and are present at significantly higher rates ($\omega>1$), we would observe evidence of adaptive evolution (Kimura, 1977)(Miyata & Yasunaga, 1980).

In the simplest models $\omega=d_N/d_S$ measures the direction and magnitude of changes on the amino acid level for a whole protein coding sequence. However, straightforward use of this model is not very effective, considering that ω averaged over all sites is rarely greater than 1. Therefore models detecting positive selection at individual sites and/or within specific lineages have been developed (Yang, 2007)(Kosakovsky Pond *et al.*, 2005). Additionally, statistical power for detection of sites under positive selection depends on the number and divergence of investigated sequences for a given gene (Yang, 2002). Some sites under positive selection may hence not be detected or assigned with the downgraded ω value.

One example of a group of genes showing signs of adaptive evolution are genes involved in sexual reproduction (Swanson & Vacquier, 2002). Experimental studies showed that reproductive genes, especially those involved in gamete recognition, undergo rapid divergence comparing to genes from non-reproductive tissues (Civetta & Singh, 1998)(Vacquier, 1998)(Singh & Kulathinal, 2000). Selection on gamete recognition genes may result in the formation of reproductive barriers and consequently lead to speciation

(Palumbi, 2008)(Ferris *et al.*, 1997)(Civetta & Singh, 1998)(Hellberg & Vacquier, 1999)(Swanson *et al.*, 2001a). Progressing genetic isolation would be manifested by reduced fitness or sterility of hybrids. These observations are especially relevant in the marine environment which harbors many broadcast spawning species. The latter release unfertilized gametes into the water column which implies that elaborate prezygotic selection mechanisms, as is typical for many terrestrial vertebrate species, are lacking and that sexual selection is narrowed down to the individual gametes. Therefore, gamete recognition proteins have been hypothesized to be under strong directional selection. The research conducted on sea urchin speciation and evolution of bindin, a protein covering the acrosomal process of the sperm and binding in a species-specific way to a receptor on the egg plasma membrane prior to the fusion of sperm and egg cells, is exemplar in this respect (e.g. (Glabe, 1979)(Metz & Palumbi, 1996)(Vacquier & Moy, 1997)(Zigler, 2008) (Levitan & Ferrell, 2006)(Lessios *et al.*, 2012).

To date, no such studies have been conducted on algal species, even though plus and minus sexual agglutinins mediating sexual adhesion between *Chlamydomonas* gametes have been characterized in detail (e.g. Adair et al. 1982; Goodenough et al. 1985; Ferris et al. 2005). Here we focus on the evolution of genes hypothesized to be involved in sexual reproduction and gamete recognition of the brown algal model *Ectocarpus siliculosus*.

Ectocarpus is a cosmopolitan genus composed of three recognized species: *E. siliculosus, E. fasciculatus* and *E. crouaniorum*. Crossing experiments have been performed between a large number of strains (e.g. (Muller, 1979a)(Muller, 1988)(Stache, 1990)(Muller & Eichenberger, 1995)(Stache Crain *et al.*, 1997)(Peters *et al.*, 2010b)) and yielded detailed observations on gametic compatibility. Crossing experiments showed complete sterility between female *E. fasciculatus* and male *E. siliculosus* and post-zygotic developmental barriers in male *E. fasciculatus* – female *E. siliculosus* hybrids as well as in crosses of the former species with *E. crouaniorum* (Muller & Eichenberger, 1995)(Peters *et al.*, 2010b). However, cases of pre-zygotic or post-zygotic sterility were frequently observed also between *E. siliculosus* strains (Muller, 1979a)(Muller, 1988)(Stache, 1990)(Stache Crain *et al.*, 1997). Phylogenetic distances between clades approximately related to the degree of strain incompatibility correlate with results from cross-fertilization experiments (Table 1)(Figure 1) and imply that more *Ectocarpus* species may exist in reality (Stache Crain *et al.*, 1997)(Peters *et al.*, 2010b). Nevertheless, it could not be established if the incompatibility is an effect of selection on mating genes, because the genetic determinants of sex and fertilization in *Ectocarpus* are still elusive.

Here we describe a molecular evolution study of nine *Ectocarpus* genes expressed specifically during reproduction. With the aid of Next Generation Sequencing we were able to obtain comprehensive transcriptome data from *E. siliculosus* gametes (Lipinska *et al.*, unpublished). Several genes with gender biased expression and potential gamete recognition function were identified. These provided a subset to test for positive selection at the amino acid level using *Ectocarpus* species of known sexual compatibility. We found polymorphism-based evidence of selective pressure acting on at least one of the investigated genes. Interestingly, that gene displayed domain similarities to the receptor for egg jelly (REJ) protein involved in sperm-egg recognition in sea urchins (Moy *et al.*, 1996a).

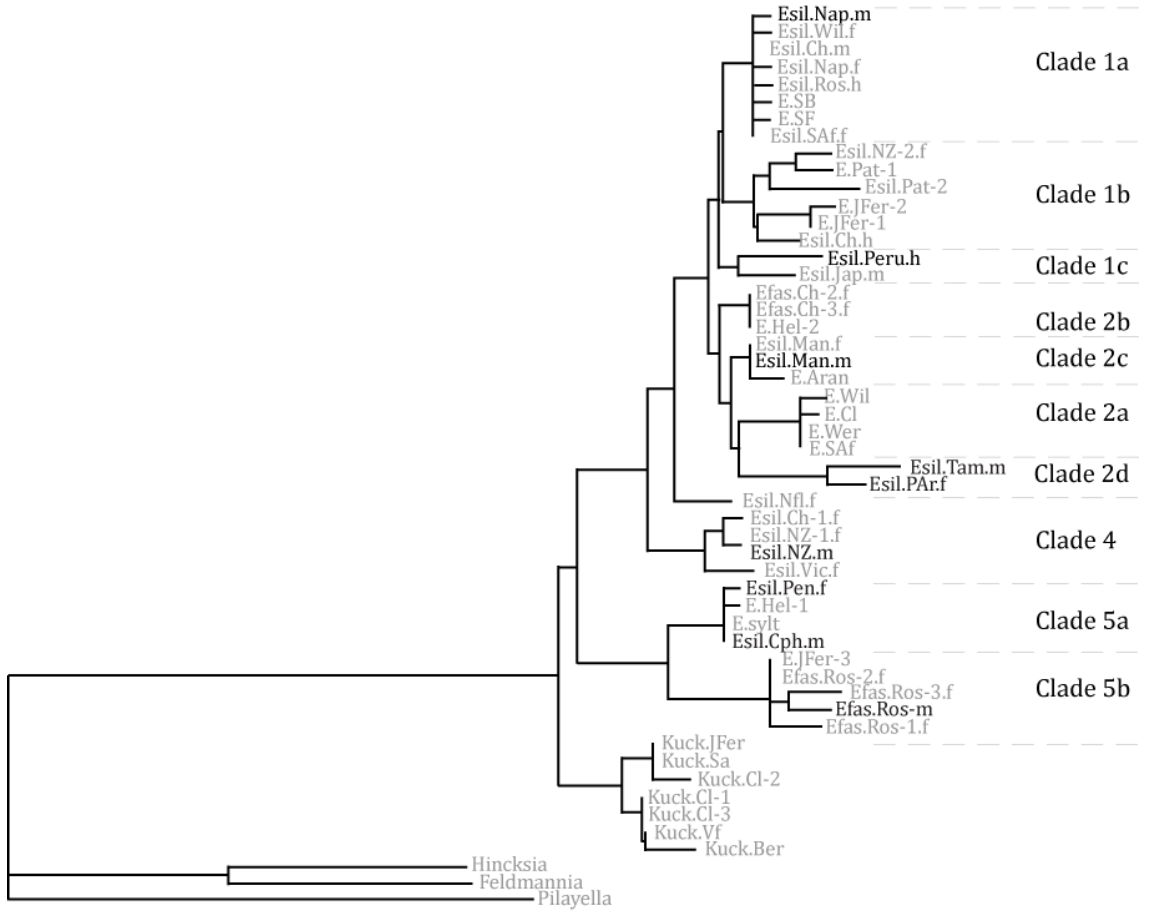


Figure 1. Neighbor-joining tree of 52 strains of *Ectocarpus* (Esil) and *Kuckuckia* (Kuck) based on Rubisco spacer with assigned clade numbers (reproduced from Stache-Crain (1997)). Strains used in this study are marked black.

Table 1

Ectocarpus cross-strain fertility

		Male								
		Nap 1a	**Per 1c**	**Cro 2c**	**Tam 2d**	**PAr 2d**	**NZ 4**	**Cph 5a**	**Pen 5a**	**Fas 5b**
Female	**Nap, 1a**	F	F	post	z	z	post	z	pre	post
	Per, 1c	F	F							
	Cro, 2c	post		F	z	z			pre	post
	Tam, 2d	z		z	F	z			pre*	
	Par, 2d	pre		z	z	F			pre*	
	NZ, 4	post					F			
	Cph, 5a	z								
	Pen, 5a	pre		pre	pre*	pre*			F	
	Fas, 5b	pre		pre						F

Data on cross-fertility summarized from (Müller, 1979) (Stache, 1990)(Muller & Eichenberger, 1995)(Stache Crain *et al.*, 1997)(Peters *et al.*, 2010b). Clade numbers corresponding to the phylogenetic position are given with the strain names. F – full interfertility, z – zygote formation, no data on growth, pre – prezygotic barriers, no cell fusion, post – hybrids with reduced growth or non-functional reproductive structures. (*) no data on the actual Penikese strain (Pen) are available; fertilization data were inferred from Woods Hole, Massachusetts strain of similar restriction in the mating pattern with other strains, but completely interfertile with the Penikese strain.

2 Materials and Methods

DNA extraction

DNA was extracted from 20-50 mg of culture material (*Ectocarpus*) or silica dried specimens (*Scytosiphon*). Samples were ground in liquid nitrogen and extracted using pre-heated CTAB buffer (100mM Tris-HCl, pH 8.0, 1.4 M NaCl, 20mM EDTA, 1% w/v CTAB, 1% w/v PVP, 0.2mg/ml proteinase K), followed by 30 minutes incubation at 60°C. After 10 minutes centrifugation at full speed, the aqueous phase was extracted with chloroform:isoamyl alcohol (24:1). DNA was precipitated with 80% isopropanol for 30 minutes on ice and next cleaned with Qiagen MagAttract Suspension G according to the manufacturer's instructions.

Sequence data collection

Taxon sampling comprised nine *Ectocarpus* strains representing clades 1a, 1c, 2c, 2d, 4, 5a and 5b as described by Stache Crain (1997). *Scytosiphon "lomentaria"* (specimens ODC1349 and LT0114 deposited in at University Gent, Belgium) was used as an outgroup (Figure 2, Table 2). Strains designated by 'CCAP1310' are maintained at the Culture Collection of Algae and Protozoa, Oban, Scotland. Isolates described by 'Ec' are housed in the macroalgal culture collection at Roscoff, France. Twelve data sets of nuclear protein-coding genes (Table 3) and the mitochondrial *cox*3 gene were analyzed. Primers were designed based on

the available *Ectocarpus* genome sequence (Cock *et al.*, 2010) using the ORCAE platform (Sterck *et al.*, 2012) (Supplementary Table 1).

PCR amplification

Primers were designed using Primer3 software (Rozen and Skaletsky, 2000) with default settings (Supplementary Table 1). The touch-down PCR reactions were performed based on the recommendations of Korbie and Mattick (2008) using listed conditions: 3 min at 95^0C, followed by phase 1: 15 cycles of 30 sec at 95^0C, 30 sec at 65^0C (-1^0C every next cycle), 60 sec at 72^0C and phase 2: 20 cycles of 30 sec at 95^0C, 30 sec at 50^0C, 60 sec at 72^0C. DNA Taq polymerase and dNTPs were obtained from Invitrogen (Life Technologies). Product size was checked on a 1% agarose gel with GeneRuler™ 1 kb DNA Ladder (Fermentas) and sequenced using the Sanger method.

Phylogenetic analyses

*Cox*3 sequences were generated for phylogenetic analysis. Obtained sequences were compared with *Ectocarpus* sequences in GenBank using blastn (Altschul *et al.*, 1997). Maximum likelihood analyses were carried out with RAxML version 7.7.1 (Stamatakis *et al.*, 2008) on the RAxML blackbox server (http://phylobench.vital-it.ch/raxml-bb/) using the CAT model (Stamatakis, 2006). Searches were started from 200 distinct randomized maximum parsimony starting trees and branch support was assessed with the classic bootstrapping algorithm (1000 replicates). Bayesian phylogenetic inference was carried out with MrBayes version 3.2 (Ronquist *et al.*, 2012). Two independent runs, each consisting of four incrementally heated chains, were run for 3 million generations using default priors and other settings. Trees were sampled every thousand generations. Convergence of likelihood and parameter values was assessed with Tracer version 1.5 (Rambaut & Drummond, 2009) and a suitable burn-in value was chosen (burnin = 500). Bayesian posterior probabilities for clades were computed from the post burn-in sample of trees and indicated on the ML tree (Figure 2).

Positive selection analyses

Codon-based nucleotide alignments were used in conjunction with the species tree (see above). Individual exon sequences representing the same gene for a given strain were concatenated using FaBox (Villesen, 2007). Input files for analyses using CODEML in the PAML 4 suite version 4.7 (Yang, 2007) were generated by first aligning nucleotide sequences using the CLUSTAL-W algorithm implemented in MEGA version 5.05 (Tamura *et*

al., 2011), converting the alignment to amino acids, adjusting manually when necessary, and transforming to the PAML 4 required format using FaBox (Villesen, 2007).

Nonsynonymous (d_N) and synonymous (d_S) rates were estimated by the maximum likelihood method available in PAML 4 using the F3X4 model of codon frequencies. ML analysis allows for a more realistic assessment of diversifying selection, as it takes into account unequal transition and transversion rates as well as unequal codon and base substitutions. Paired nested site models (M0, M3; M1a, M2a; M7, M8) (Yang, 1998)(Yang *et al.*, 2000a)(Yang, 2000) of sequence evolution implemented in CODEML were used in this analysis. The Second model in each pair (M3, M2a and M8) is derived from the first by allowing variable ω ratios between sites, making it possible to detect positive selection at critical amino acid residues. Importantly, these methods can be used without prior knowledge of present protein domains or molecular activity and may help to identify functionally important regions. Next nested models were compared using the likelihood ratio test (LRT). LRT is used to justify if the additional parameters incorporated in the more complex model (allowing sites under positive selection) are making a statistically significant improvement in fit to the data. The LRT is represented by twice the difference in maximum likelihood scores for the two models under comparison and determining the chi-square critical value assuming either 4 degrees (M0, M3) or 2 degrees of freedom (M1a,M2a; M7, M8). Empirical Bayes methods allowed for identification of positively selected sites a posteriori (Yang *et al.*, 2000a)(Yang, 2007).

Since the signature of positive selection might be overlooked due to its transient or episodic nature, we performed, in parallel, an analysis of occasional selection at individual sites using MEME (Mixed Effects Model of Evolution), implemented in HyPhy (Murrell *et al.*, 2012) available at Datamonkey.org server (Delport *et al.*, 2010)(Kosakovsky Pond *et al.*, 2005) and compared it with the results from PAML 4. MEME allows the distribution of ω to vary between sites and branches, assuming that ω is not necessarily constant over time and purifying selection in some lineages may mask the signal of positive selection in others (Murrell *et al.*, 2012). The F81 codon substitution model was used in HyPhy (Felsenstein, 1981).

Table 2

Strains used in this study

Lineage	Strain code or species	Short name	Strain no.		Origin
			Ec*	CCAP**	
5a	Ec sil Cph 40-11 m	Cph	-	CCAP 1310/100	Copenhagen, Denmark
2d	Ec sil Tam 2b m	Tam	-	CCAP 1310/122	Tampa, Florida
5a	Ec sil Pen 2a m	Pen	-	CCAP 1310/111	Penikese Island, Massachusetts
2d	Ec sil PAr 18a m	Par	-	CCAP 1310/108	Port Aransas, Texas
5b	E. fasciculatus	Fas	Ec185	-	Perharidy, France
1a	E. siliculosus	Nap	Ec400	CCAP 1310/329	Naples, Italy
2c	E. crouaniorum	Cro	Ec477	-	Perharidy, France
1c	E.siliculosus (genome strain)	Per	Ec32	CCAP1310/4	San Juan de Marcona, Peru
4	Ec sil NZKU 1-3 m	Nze	-	CCAP 1310/56	Kaikoura, New Zealand
-	Scytosiphon "lomentaria"	-	ODC1349***		Brittany, France
-	Scytosiphon "lomentaria"	-	LT0114***		Australia

Protein functional characteristics prediction

Protein structure prediction was done using Phyre2 (Protein Homology/Analogy Recognition Engine) (Kelley & Sternberg, 2009). Transmembrane helices and their topology were inferred from memsat-svm implemented in Phyre2 or from the TMHMM server v. 2.0 (Krogh *et al.*, 2001)(Sonnhammer *et al.*, 1998). To determine functional domains we performed a Gene Ontology (GO) and protein domain search (InterPro database) using Blast2GO v. 2.6.6 (Conesa *et al.*, 2005) with E-value Hit Filter set to 1.0e-6. Information about the signal peptides was derived from *Ectocarpus* gene description pages at ORCAE (Sterck *et al.*, 2012).

Table 3

Genes of interest used for the evolutionary analysis study

Gene	Sequence description	Length (bp)*	Signal peptide	TM domains	GO IDs	InterPro domains
Esi0146_0035	Hypothetical protein	2817/ 1311	Y	1 TM	-	-
Esi0130_0068	Conserved unknown protein	4284/ 417	N	7 TM	GO:0005515 protein binding GO:0007218 neuropeptide signaling pathway GO:0016020 membrane	IPR000203 GPS domain IPR002859 PKD/REJ-like protein domain IPR003018 GAF domain
Esi0008_0013	Conserved unknown protein	6693/ 195	N	6 TM	GO:0007218 neuropeptide signaling pathway GO:0016020 membrane GO:0046872 metal ion binding GO:0016021 integral to membrane GO:0016567 protein ubiquitination GO:0008270 zinc ion binding GO:0004842 ubiquitin-protein ligase activity GO:0007155 cell adhesion	IPR002859 PKD/REJ-like protein
Esi0214_0046	Polymorphic outermembrane protein	4644/ 821	Y	8 TM	GO:0030154 cell differentiation GO:0016020 membrane	IPR006626 Parallel beta-helix repeat IPR011050 Pectin lyase fold/virulence factor IPR012334 Pectin lyase fold
Esi0123_0056	Similar to G-protein coupled receptors	1650/ 175	Y	8 TM	GO:0004872 receptor activity GO:0004672 protein kinase activity GO:0016021 integral to membrane	IPR009637 Transmembrane receptor, eukaryota
Esi0163_0051	conserved unknown protein	3918/ 117	N	2 TM	-	-
Esi0160_0060	conserved unknown protein	1782/ 186	N	extracellular	GO:0016020 membrane GO:0007218 neuropeptide signaling pathway	IPR002859 PKD/REJ-like protein IPR014010 Egg jelly receptor, REJ-like
Esi0050_0107	conserved unknown protein	3381/ 402	N	6 TM	GO:0007218 neuropeptide signaling pathway GO:0016020 membrane	IPR000203 GPS domain IPR002859 PKD/REJ-like protein IPR014010 Egg jelly receptor, REJ-like
Esi0101_0018	Tubular mastigoneme-related protein, Sig1	2184/ 544	Y	extracellular	GO:0005515 protein binding	IPR000742 Epidermal growth factor-like domain IPR013032 EGF-like, conserved site

*Length of a whole coding sequence and a gene part sequenced in this experiment (bp).
Transmembrane domains (TM) are predicted by TMHMM algorithm (Sonnhammer *et al.*, 1998), Gene Ontology (GO) categories and protein domains are annotated using Blast2GO (Conesa *et al.*, 2005).

3 Results and Discussion

3.1 NextGen sequencing screening for candidate reproductive genes

We constructed the whole transcriptome library for male and female gametes of *Ectocarpus siliculosus* using Next Generation Sequencing platform SOLiD 3 (Lipinska et al., unpublished). To enrich the library for the gamete specific transcripts we performed a differential expression study as well as comparative analyses with available EST data from sporophyte and gametophyte tissues. Almost one third of the gamete transcripts were allocated exclusively to the reproductive cells with a significant enrichment of signal transduction gene ontologies (Lipinska et al., unpublished). We identified 563 female and 967 male genes with gender specific expression which were investigated further for the presence of signal peptides, transmembrane helixes and functional domains potentially involved in cell-cell recognition. Fertilization experiments in *Ectocarpus* have shown the presence of lectin-like receptors in male gametes and compatible carbohydrate moieties presented on female gamete surface (Schmid, 1993). Therefore, we restricted the search to male expressed genes coding for extracellular or cell surface proteins with potential receptor activity. Since signal peptide prediction relies on a correct assignment of the first exon and some proteins contain internal targeting sequence, we focused mainly on the presence of transmembrane domains as a criterion of subcellular localization.

We selected nine genes of interest (Table 3) and three house-keeping genes (actin-related protein 2 Esi0289_0026, mannuronan C-5-epimerase Esi0069_0059 and heat shock protein 90 Esi0138_0009) for sequencing in nine representative strains of *Ectocarpus* and *Scytosiphon lomentaria* (Figure 2, Table 2). We obtained sequence reads representing coding regions from one or few exons of a joined length >100bp.

In four of the genes a REJ-like (Receptor for Egg Jelly) domain (IPR002859) was predicted. REJ-like domains are found in the sperm proteins of sea urchins as well as the human PKD1 (Polycystic Kidney Disease 1) protein. The exact function of the REJ domain is unknown. However, the unique localization and expression pattern of REJ proteins in sea urchins and humans suggest their central role in fertilization (Moy *et al.*, 1996a)(Hughes, 1999). Sea urchin sperm contains several members of the REJ protein family (named SpREJ1-10), out of which SpREJ1 binds a fucose sulfate polymer of the egg jelly, triggering the acrosome reaction and transforming the sperm into a fusogenic cell (Moy *et al.*, 1996a)(Vacquier & Moy, 1997). Some members of the SpREJ and human PKD families share homology with cation channels and almost all of them possess a G-protein coupled receptor cleavage site (GPS) upstream of the first transmembrane helix (Gunaratne *et al.*, 2007). This topology was found in

Esi0130_0068 and Esi0050_0107; whereas Esi0008_0013 and Esi0160_0060 showed partial similarity to the polycystin cation channel protein (TIGR00864) but revealed no GPS domain (Pfam01825).

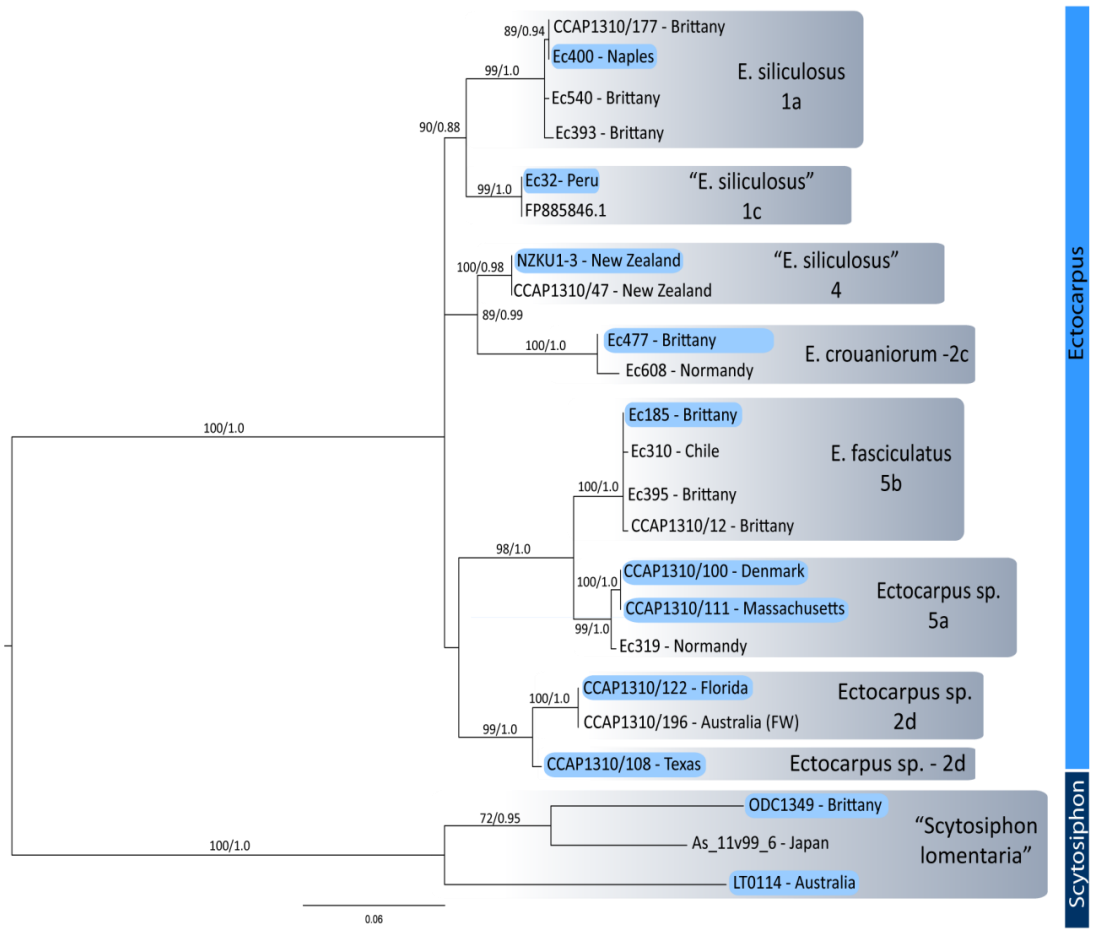


Figure 2. Maximum likelihood tree based on the *cox*3 gene from nine *Ectocarpus* strains and two *Scytosiphon "lomentaria"* specimens. Strains used in this study are indicated with blue bars. ML bootstrap (left) and Bayesian posterior probability (right) values are shown. The strains from New Zealand (used to generate gamete transcriptome data) and Peru (genome sequenced strain) are referred to as *E. siliculosus*, however their affiliation is uncertain.

Substantial evidence supports a role of G-protein coupled receptors (GPCRs) in egg-sperm interactions during fertilization. For example, a GPCR located on the spermatozoa plasma membrane activates a signaling pathway responsible for the zona pellucida induced acrosomal exocytosis in mouse (Ward *et al.*, 1992). Human PKDREJ was found to modulate G-protein signaling (Sutton *et al.*, 2006) and a G-protein coupled cAMP transduction pathway

was implicated in chemotaxis in human sperm (Spehr *et al.*, 2004). Therefore, one of the male expressed GPCR-like receptors (Esi0123_0056) was included in the analysis.
The additional genes consisted of a polymorphic outer membrane protein (Esi0214_0046), a sexually induced gene 1 (Sig1) (Esi0101_0018) and two male specific unknown proteins (Esi0146_0035, Esi0163_0051). Sig1 belongs to the family of Stramenopile mastigoneme genes, expressed during reproduction but of unresolved function (Armbrust, 1999)(Honda *et al.*, 2007)(Yamagishi *et al.*, 2009). However, striking diversification of Sig1 between closely related species of a diatom *Thalassiosira* (Armbrust & Galindo, 2001) as well as some evidence of positive selection (Sorhannus & Kosakovsky Pond, 2006), led to the speculation of its role in gamete recognition.

3.2 Positive selection analysis

To search for evidence of adaptive evolution in gamete proteins, we performed maximum likelihood analysis implemented in PAML 4 (Yang, 2007). To give an indication of the sequence divergence level and the selective constraint acting upon each gene, we provide the transition/transversion ratio (κ), the average nonsynonymous/synonymous rate (ω), and the tree length estimated under the simple model M0 (one ω ratio for all sites)(Goldman & Yang, 1994) (Table 4).

Although only one gene (Esi0130_0068) showed statistical evidence for adaptive evolution in the PAML 4 analysis, the proportion of positively selected genes might be underestimated due to the small taxon number and incomplete cds sequences used. It is worth noting that four other analyzed genes (Esi0050_0107, Esi0146_0035, Esi0163_0051 and Esi0214_0046) displayed overall ω values ≥0.3. As it was shown in *Drosophila* (Swanson *et al.*, 2004), genes with a d_N/d_S ratio >0.3 were still very likely to have been subjected to adaptive evolution and are thus good candidates for further investigation. Additionally, we found evidence of significant variation ($p<0.05$) in the d_N/d_S ratio for three out of four of these genes (except for Esi0163_0051) using the discrete model M3. Nevertheless, model M3 is not a robust test of positive selection and should be considered only as an indication of possible candidates for further evaluation. The MEME analysis in HyPhy identified positively selected sites in Esi0050_0107, Esi0146_0035 and Esi0214_0046 as well as in Esi0130_0068. However, the latter presented different sites under selection depending on the model used (Table 5).

Table 4

Parameter estimates under model M0 (one ratio) in PAML 4

Gene	lnL	ω=dn/ds	κ	tree length
Esi0008_0013	-425.543781	0.12086	2.06379	0.54180
Esi0050_0107	-1043.678828	0.39362	2.89208	0.98168
Esi0101_0018	-1032.771895	0.05217	3.05085	0.50913
Esi0123_0056	-287.087423	0.09287	8.37300	0.28151
Esi0130_0068	-1187.890945	0.84501	1.60804	1.07654
Esi0146_0035	-3200.128608	0.40455	2.67169	0.76487
Esi0160_0060	-326.363395	0.12243	3.59600	0.34210
Esi0163_0051	-252.032792	0.41725	1.91659	0.63725
Esi0214_0046	-2150.400176	0.29229	2.25638	1.10022
Esi0289_0026	-808.882093	0.00010	4.12690	0.28558
Esi0069_0059	-337.408019	0.01173	8.59597	0.64678
Esi0138_0009	-262.252627	0.03101	1.47486	1.46131

Table 5

Positively selected sites by the site-prediction methods in PAML 4 and HyPhy (DATAMONKEY)

		Pr>90%		p value <0,1
Gene	$\omega=d_N/d_S$ *	M1a-M2a	M7-M8	MEME
Esi0008_0013	0.1	none	none	none
Esi0050_0107	0.4	none	none	88, 184
Esi0101_0018	0.1	none	none	none
Esi0123_0056	0.1	none	none	none
Esi0130_0068	0.8	293, 297	293, 297, 302	214, 224, 230, 232,
Esi0146_0035	0.4	none	none	374, 393, 397
Esi0160_0060	0.1	none	none	none
Esi0163_0051	0.4	none	none	none
Esi0214_0046	0.3	none	none	1062, 1063, 1086, 1280
Esi0298_0026	0.0	none	none	none
Esi0069_0059	0.0	none	none	none
Esi0138_0009	0.0	none	none	none

*estimate of dN/dS assuming no rate heterogeneity

3.3 Intraspecies polymorphism and evidence of positive selection in Esil0130_0068

Esi0130_0068 has 1427 amino acids and is presumed to possess 7 (TMHMM algorithm) or 6 (memsat-svm algorithm) transmembrane domains. The N-terminal region, coded by 976 amino acids, is predicted to be extracellular and this segment was targeted for resequencing in representative *Ectocarpus* strains. Blast searches identified a GPS domain and a REJ domain with E-values of 3,78e-05 and 3,98e-30 respectively. Inside the REJ region, a

polycistin cation channel (PCC), a Na^+, K^+, Ca^{2+} channel activated by Ca^{2+}, is also found with an E-value of 3,01e-05.

All models allowing individual sites to evolve under positive selection (M3, M2a, M8) gave a significantly better fit to the Esi0130_0068 data (Table 6) and identified a substantial proportion of sites with d_N/d_S >1 (Table 7, Figure 3). This result is consistent with an evolutionary history characterized by frequent episodes of positive selection. Model M3 suggests ~13% of sites under positive selection with ω2=3,49. Similarly, model M2a found 12% sites under positive selection with ω2=3,59 and model M8 suggests 12% of sites with ω1=3,55. The LRT (2ΔL) statistics for comparison of M1a-M2a models equals 7,35 (p=0,0253) and for M7-M8 models is found to be 7,55 (p=0,0229) compared with χ^2 distribution with d.f. = 2. The codons inferred to be under positive selection by PAML 4 with posterior probability >90% lie within the REJ domain in the extracellular region of the protein. Sites identified by HyPhy are adjacent to the N-terminal site of the REJ domain (Figure 4) and are also indicated by the Empirical Bayes analysis in PAML 4, however below the significance threshold. Noteworthy, several positively selected sites in the REJ domain and its flanking extracellular regions were detected (PAML model M8) when the sequence of sperm PKDREJ in primates was investigated (Hamm *et al.*, 2007).

Table 6

Likelihood ratio statistics (2delta L)for Esi0130_0068

Comparison	2delL	df	chi-squared 5%	chi-squared 1%
M0 (one ratio) vs. M3 (discrete)	20,544212	4	3,753	13,28
M1a (nearly neutral) vs. M2a (positive selection)	7,350578	2	5,99	9,21
M7 (beta) vs. M8 (beta&w)	7,551876	2	5,99	9,21

Table 7

Parameter estimates and log-likelihood values under models of variable w ratios among sites for Esi0130_0068

Model	p	Parameters	*l*	d_N/d_S	Positively selected sites BEB (Pr>90%)
M0: one ratio	1	ω=0,84501	-1187,89	=ω	None
M1a: nearly neutral	1	p0=0,36446, ω0=0,00000, p1=0,63554, ω1=1,00000	-1181,3	0,6355	Not allowed
M2a: positive selection	3	p0=0,38710, ω0=0,05747, p1=0,48892, ω1=1,00000, p2=0,12399, ω2=3,59092	-1177,62	0,9564	42, 46
M3: discrete	5	p0=0,31976, ω0=0,0000, p1=0,54469, ω1=0,88415, p2=0,13555, ω2=3,49354	-1177,62	0,9551	42, 46
M7: beta	2	p = 0,02165 q =0,01304	-1181,39	0,6035	Not allowed
M8: beta&w	4	p0 = 0,87174 p = 0,10745 q = 0,07982, (p1 = 0,12826) ω = 3,55267	-1177,62	0,9559	42, 46, 51

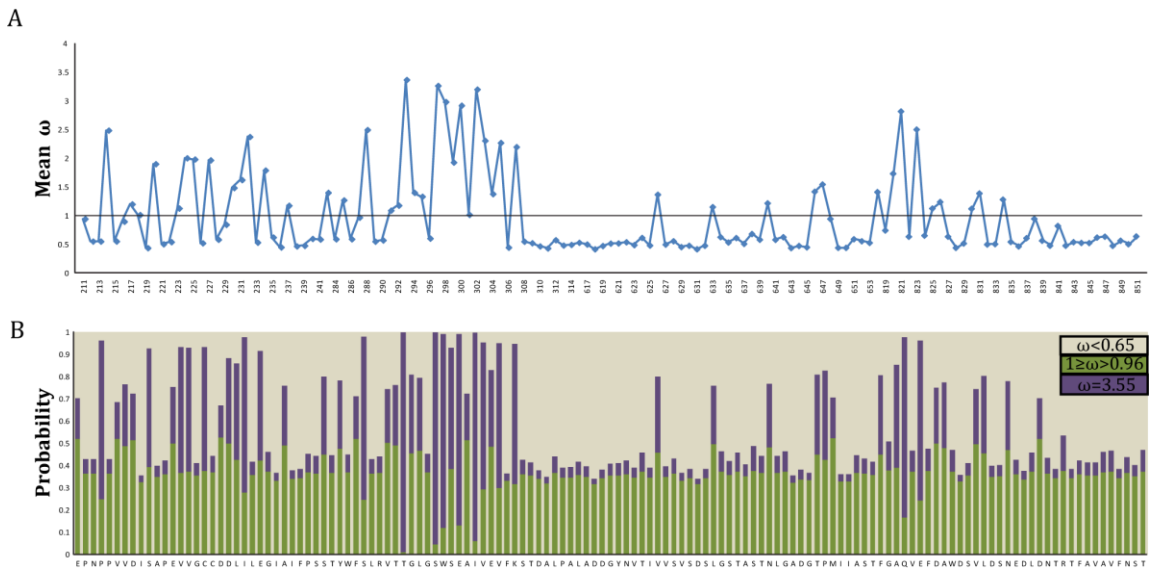


Figure 3. A) Posterior probabilities for site classes in Esi0130_0068 as calculated by M8 (beta&ω) model in PAML. The ω ratios are 0.00000, 0.00005, 0.00618, 0.12717, 0.64633, 0.96054, 0.99825, 0.99997, 1.00000, 1.00000 and 3.55267. Each of the first 10 categories has a proportion of 0.08717 and the last category has a proportion of 0.12826. Categories are grouped by ω<0.65, 1≥ω>0.96, ω=3.55. B) Posterior means of ω, calculated as the average over 11 site classes. The amino acid sequence represents the genome strain (Ec32), the site numbers correspond to the position in the coding sequence.

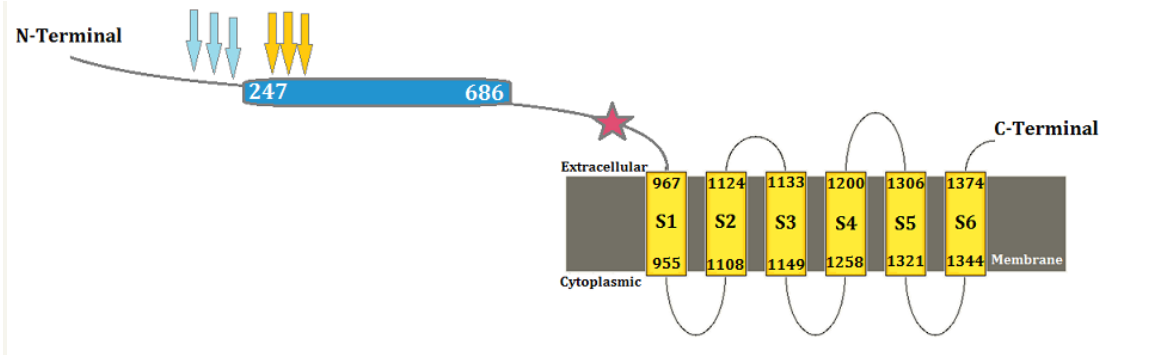


Figure 4. Topology of Esi0130_0068 as predicted by Phyre2 and Blast2GO.

The extracellular and cytoplasmic sides of the membrane are labeled and the beginning and end of each transmembrane helix are indicated by the residue index. GPS domain is indicated by a red star and REJ domain is presented as a blue bar. Positively selected sites identified by PAML 4 are marked with orange arrows, sites identified by HyPhy are marked blue.

No specific protein domains were associated with the positively selected sites indicated by MEME in Esi0146_0035 (sites located to an extracellular region) and Esi0214_0046 (sites located to an intracellular loop). However, sites in Esi0050_0107 belonged to a region recognized by the Phyre2 engine (Kelley & Sternberg, 2009) as a carbohydrate-binding module (endoglucanase) with high confidence (>97% confidence, sequence identity 13%, amino acid residues 40-298). At this level of confidence, it is highly probable that the protein of interest adopts the overall fold predicted, even if the sequence identities are below 20% (Kelley & Sternberg, 2009).

4 Conclusion

Maximum likelihood analyses of *Ectocarpus* genes expressed specifically in male gametes revealed significant variation in selective pressure among sites in one of the investigated proteins. The analysis by HyPhy also suggests that there are possible cases of episodic adaptive evolution (p value <0.1) in three other genes studied here (Esi0214_0046, Esi0050_0107 and Esi0146_0035).

Adaptive evolution has been commonly observed in reproductive proteins (Clark *et al.*, 2006). Examples include *Drosophila* species where fertilization and mating traits showed a lack of evolutionary constraints, suggesting that they were shaped during the early stages of speciation by directional sexual selection (Civetta & Singh, 1998). A particularly large proportion of positively selected sites are found in the sperm-egg binding moieties (Swanson & Vacquier, 1998)(Swanson *et al.*, 2001b)(Clark *et al.*, 2009)(Mah *et al.*, 2005), indicating that these interactions might be subjected to selective pressure. In general, male biased genes show a faster rate of evolution than female biased genes, but they also show a narrower expression range (Zhang *et al.*, 2004)(Assis *et al.*, 2012)(Parsch & Ellegren, 2013). Typically, such limited expression pattern and evolutionary rate comparisons define sex-biased genes. The classes of genes analyzed in this study represent genes with evidence of expression defined only to male gametes of *Ectocarpus* and possibly subjected to adaptive evolution. By interpolation, the observed positive selection may pinpoint the genes that are directly involved in male reproduction, which would be an important step towards understanding the molecular basis of gamete interaction during reproduction in *Ectocarpus*. Interestingly, Esi0130_0068 resembles the topology of the egg recognition protein in sea urchin sperm. The presence of a REJ domain in combination with a GPS motif and possible cation channel function makes it an appealing candidate for *in situ* evaluation.

Since gamete recognition is vital for successful reproduction, changes in receptors' compatibility may have implications in the establishment of reproductive barriers and the

speciation process (Palumbi, 2008). Molecular evolution of genes involved in gamete recognition in marine invertebrates revealed several amino acid sites under strong diversifying selection (Lee *et al.*, 1995)(Metz *et al.*, 1998)(Hellberg & Vacquier, 1999)(Mah *et al.*, 2005)(Zigler, 2008). In addition, variable selective pressure exerted on the lysin gene in closely related (ω >1) and more diverged species (ω <1) of abalone suggested, that diversifying selection acts on closely related sympatric species, whereas distantly related species are already relieved from it (Lee *et al.*, 1995)(Yang *et al.*, 2000b). This could be an example of reinforcement (Dobzhansky, 1951) where sex involved genes would be under selective pressure to establish barriers to reproduction in reunited populations. In fact, sites identified as being influenced by positive selection are frequently associated with protein domains thought to mediate recognition between sperm and eggs (eg. (Swanson *et al.*, 2001b)). Proteins responsible for this type of interactions in brown algae are not discovered yet; however a Sexually induced gene 1 (Sig1) was hypothesized to play a role in gamete adhesion in diatom *Thalassiosira* (Armbrust, 1999)(Armbrust & Galindo, 2001). *Ectocarpus* possesses one homolog of Sig1 with similar length and domain architecture with four EGF domains (Epidermal growth factor, IPR000742), three in close proximity to the N-terminal and one at the C-terminal protein end. Similarity of these domains to cell adhesion and signaling factors as predicted by Phyre2 (Kelley & Sternberg, 2009) with confidence >99% (seq. identity 32-18%) advocates for its role in cell-cell interaction. Additionally, Sig1 in *Thalassiosira* showed high nonsynonymous sequence divergence (within the first 250 amino acids) between closely related species (Armbrust & Galindo, 2001). In this study, only exons 5 and 6 corresponding to the *Ectocarpus* Sig1 (Esi0101_0018) C-terminal sequence (amino acids 514-721) were analyzed and no sites under positive selection were detected. The remaining exons 1-4, that might bear the positively evolving sites, could not be amplified by PCR from the genomic DNA and are thus missing in this analysis.

It has been shown, that episodic selection is widespread and codon sites interfering with the protein function not always show high ω values, thus the number of sites experiencing positive selection may have been generally underestimated (Nozawa *et al.*, 2009)(Murrell *et al.*, 2012). Nucleotide sequences used in this study represent only a partial coding sequence of selected genes in a limited number of strains. More conclusive evidence of directional selection and forces shaping the divergence in the sex-related genes in *Ectocarpus* would require a larger sample size and complete gene sequences for a better estimation of evolution over time. In addition, some form of experimental confirmation would be necessary to fully understand the adaptive changes and their implications.

Supplementary Table1

Primers used for polymerase chain reaction

Gene	Primer name	Sequence 5'-3'
cox3	cox3P1F	GAYCCWAGTCCMTGGCCWTTAG
	cox3P2R	ACAAARTGCCAATACCAAGC
Esi0289_0026	289_26F	CTGAACCCAACCAAGAACCG
	289_26R	TTGATGTCCTTCTCCAGCCG
Esi0069_0059	69_59F	GAGATGCAACAACGTCGAGA
	69_59R	TCGAACGTGTTGTTGGTGAT
Esi0101_0018	101_18_5F	GACACCGACGGAGAGTTC
	101_18_5R	GTAGGCGTTGTAGTCCTCTA
	101_18_6F	AGATCAAGCTGGACAGGC
	101_18_6F	TGTGTATCGCAGTTCTCATT
Esi0130_0068	130_68_5F	ATCGGGGCCTTTCTCTCC
	130_68_5R	TGAAGGGAAGATCGCGATTC
	130_68_7F	CCGGCGCTCAAAATGGGTT
	130_68_7R	ACGCGGGAAGGGCATCAGT
	130_68_13F	GGTTGTCTTCATCTTGGTCA
	130_68_13R	GAGGAGTAGATGCGATGATC
	130_68_17F	CAGAGCACATCCGATTGAT
	130_68_17R	TGTGGAATTGAATACTGCCA
Esi0138_0009	HSP90F	GGA TGA AGA TCG ACG AGT CC
	HSP90R	TCA TGA GCT TGC ACA GAC C
Esi0146_0035	146_35_1F	AATCCGATTCTTGCTAGCGG
	146_35_1R	CTGGATGCCTGATGTGGGA
	146_35_3F	GTGTGTGATTATCGTTGCGGA
	146_35_3R	TTCATCATTGGCAGCAGACG
	146_35_4F	GACACCTTCCTAACGTCACC
	146_35_4R	CGTCGGCCAAGTACTCGTAT
	146_35_5F	CTATTCCCGCAGTCCACCT
	146_35_5R	ATGACACGAATGGGAAACGG
Esi0123_0056	123_56_3F	ACAAGAGATCAACTGCCCGA
	123_56_3R	CTGGGAGCAGCAGAGAGTC
Esi0160_0060	160_60_9F	GTACGAGTGGTGGTGTTC
	160_60_9R	CTGTATAAGTCGGTAGAATCGA
Esi0163_0051	163_51_7F	TGTGACACCATCCCGACC
	163_51_7R	TAAGCTTGACGACTCCTCCC
Esi0214_0046	214_46_6F	CTGTGTGTACCGAGGGCTT
	214_46_6R	CGCTTCGCAATCCAATACGT
	214_46_8F	GTACAAGGTCTACGCTGGGG
	214_46_8R	ACTTGTGCTGCCGTGTTTG
Esi0008_0013	8_13_8F	AGGACCTAGTGTTGGACGTG
	8_13_8R	GGTGGCCGTGATGTTGAAAT
	8_13_9F	TGAGGTTTTGGTGTCTGCTG

	8_13_9R	CATACTCCACCCCGATCTCC
Esi0050_0107	50_107_1F	GGACGAGTCTTGCATTGTCC
	50_107_1R	CCGTGTCAAAAGAACCCAGG
	50_107_3F	GTCAGTGGTAGTGGTGGGTC
	50_107_3R	TTGCGTGGAATCGGTACATG
	50_107_6F	TCGTGGAGATTATCGTGGGA
	50_107_6R	CCTCGTCAAGTGATTCGCTG

Chapter 6

General discussion

The duality in sexual reproduction, being both "the masterpiece of nature" as well as "the queen of problems in evolutionary biology" was eloquently highlighted by Graham Bell (1982). Bell noted that problems and outstanding questions related to sexual reproduction are rooted deeply at the heart of biology, yet are difficult to see due to their prevalence in nature. More than 150 years after Darwin and Mendel many questions in the field of evolution and genetics of sexual reproduction still await fully satisfactory answers, the most pronounced being: Why do organisms engage in sex? Why and how does it happen in the form we observe it, with separation of sexes, various degrees of sexual dimorphism, and transition of generations?

In this thesis we explored a small fragment of this enigmatic domain of biology using the isogamous brown alga *Ectocarpus siliculosus*. We examined the biochemistry of cell-cell recognition during reproduction (Chapter 2), we used a new, cutting edge, sequencing technology (Chapter 3) to uncover gamete differentiation at the molecular level (Chapter 4), and studied evolutionary adaptation of sex-biased genes (Chapter 5). We uncovered profound differentiation in transcriptomes of morphologically indistinguishable gametes, which relates to diverged functional characteristics of male and female reproductive cells, especially in areas of carbohydrate and lipid metabolism, cellular transport, ion flux, movement, gene expression and signaling. In the latter group we described many gamete specific pathways and genes with potential function in fertilization, with some of the presumed receptors being under diversifying selection within *Ectocarpus* species. These findings make contribution to extensively studied fields of evolution of sexual reproduction, sexual selection, dimorphism and sex-linked speciation, as well as to brown algal biology.

Below, we discuss the offerings of algal models to sexual reproduction research and the relevance of genetic and genomic tools for these systems. We summarize the molecular basis for establishment of sexual roles in *Ectocarpus* gametes and the implications of adaptive changes in sex-biased genes. Finally, we highlight the interesting properties of algal lectins and their role in gamete recognition.

Algae in genomics era

After almost 50 years since the first genome sequence of a bacteriophage MS2 was determined (Fiers *et al.*, 1976), we find ourselves in the so-called post-genomics era. Technological progress has dramatically reduced the costs of sequencing, making genome projects fairly routine. The magnitude of possibilities presented by gene sequencing platforms and an ever-growing amount of accessible genomic data is now inspiring a new understanding of living organisms, acknowledging the complexity of species and genera and their inter-relationships.

Although, brown algae comprise an important eukaryotic group, such comprehensive genetic knowledge is still under development. Major progress in algal research was marked by completion of the genome sequence of *Ectocarpus siliculosus* (Cock *et al.*, 2010). Genetic analysis of algae will not only shed light on the complex environment they inhabit, but promises new insights into fundamental biological processes such as cell division, differentiation, development and reproduction, as well as phylogenetic and evolutionary analysis (Cock *et al.*, 2011)(Coelho *et al.*, 2013).

Numerous genetic, genomic and biochemical tools which are commonly applied in plant and animal research are relevant to this relatively unexplored branch of eukaryotes. For instance, mutant screens and microarray gene expression analysis were successfully used in life cycle studies in *Ectocarpus*, allowing for the discovery of molecular mechanisms involved in transition between sporophyte-gametophyte generations (Coelho *et al.*, 2011)(Arun *et al.*, 2013).

We chose the novel RNA-Seq technology to investigate the transcriptomes of *Ectocarpus* gametes. This approach allowed us to construct a genome-scale transcription map with levels of expression for each gene (Chapter 3). Unlike hybridization-based approaches, RNA-Seq is not limited to detecting transcripts that correspond to known expressed gene sequence. This was particularly important, considering that the *Ectocarpus* microarray was constructed based on ESTs from gametophytic and sporophytic tissues. A large set of gamete expressed genes overlapped with that of somatic tissue; however the whole transcriptome sequencing permitted for discovery of gamete specific transcripts comprising as much as 30% of the whole library (Chapter 4). A second advantage of RNA-Seq is its large dynamic range of expression levels for the detection of transcripts, allowing for low or very highly expressed genes to be identified. However, certain challenges of this technology have to be kept in mind still while analyzing the sequencing data.

Pre-amplified cDNA libraries often result in identical short reads, which could reflect either abundant RNA species or PCR artifacts (Wang *et al.*, 2009). Biological replicates can help to distinguish between the two scenarios; however due to low RNA availability and financial limitations these were not performed for this study. Instead, we applied a minimal number of non-overlapping (non-identical) reads per gene in order for it be considered expressed. Secondly, reads alignments are complicated for large transcriptomes because a significant portion of sequences match multiple locations in the genome. One solution consists generating longer reads, which was not available for SOLiD technology at the time when this study was conducted. We therefore used only uniquely mapped sequences for determination of the transcription levels, which might have caused underestimation of some expression

values. Finally, RNA-Seq reflects a very complex, dynamic transcriptional state of a cell. Interpretation of the data depends highly on the downstream processing methods, like stringency of the alignment quality parameters or determinants for gene expression verdict. Taken together, while RNA sequencing is powerful, experimental validation will always be necessary to confirm most results.

Algae as models in comparative genomics

Collation of genomic data from different closely to distantly related organisms can lead to the discovery of essential, conserved genes and shed light on the fundamental events in cell biology. In spite of the many cellular mechanisms common to fertilization and zygote development in almost all eukaryotes, the genes underlying these processes are largely unknown and leave room for exciting new discoveries. Only recently, a comparative transcriptome analyses of reproductive cells of the isogamous unicellular green alga *Chlamydomonas* with the malaria parasite *Plasmodium* and a higher plant (*Arabidopsis*) led to the discovery of the conserved gamete membrane fusogen HAP2(GCS1) (Mori *et al.*, 2006)(Hirai *et al.*, 2008)(Liu *et al.*, 2010)(Wong & Johnson, 2010) and a nuclear envelope protein (GEX1) necessary for transition of malaria by mosquitos (Ning *et al.*, 2013). These observations influence not only our understanding of the evolution of core reproductive processes conserved across different taxa, but could offer solutions for reproductive therapeutics or malaria treatment. Thus, comparative genomic approach using broad range of species from algae to humans offers a valuable resource to the research community in many scientific disciplines.

The independent evolution of multicellularity and reproductive systems as well as remarkable characteristics of their cell biology borrowing from both land plants and animals, as shown on the example of cytokinesis (Nagasato & Motomura, 2002), make brown algae appealing models from a comparative genomic point of view. Furthermore, the wealth of life cycles displayed by this group, with fertilization modes ranging from isogamy to oogamy, various degrees of sexual dimorphism and sex determination is particularly unique and promises advancement in many aspects of sexual reproduction studies.

Brown algae and the evolution of anisogamy

Gamete size dimorphism evolved many times independently throughout the eukaryotic tree of life, with oogamy being a nearly universal strategy for gamete production in multicellular species. The emergence of genetically differentiated sex chromosomes or sex-determining regions (SDR) is closely linked to gamete differentiation and has recently been studied in volvocine algae (isogamous *Chlamydomonas reinhardtii* versus oogamous *Volvox cartierii*)

(Ferris *et al.*, 2010)(Umen, 2011). However, the expansion of dimorphic mating loci in *Volvox* reflects not only the adaptation of oogamy and a more advanced reproduction program, but is also the singular genomic region that is substantially diverged from unicellular *Chlamydomonas*. Umen (2011) therefore concluded that evolution of oogamy may be closely linked to the acquisition of multicellularity in this lineage, whereas *Ectocarpus* combines multicellularity with a primitive isogamous sexual system.

It is generally accepted that the evolution of morphologically indistinguishable mating types has preceded the evolution of anisogamy. Since *Ectocarpus* gametophytes and gametes do not display legible dimorphism, knowledge about its sex-determining region could offer a better understanding to the subject of particular co-evolution of SDRs with the isogamy-oogamy transition and phenotypic differentiation between males and females in multicellular organisms. Future projects oriented towards comparative analyses of sex-biased gene expression in *Ectocarpus* gametes with other brown algal species with more advanced sexual dimorphism (anisogamy, oogamy) can provide information about regulatory genes involved in the emergence of these sexual systems and their evolution not only in algae, but in all living organisms. This knowledge cannot be easily gained using plant or animal models due to scarce gamete material and lack of appropriate comparative systems, emphasizing the suitability of algae in the evolution of sexual reproduction studies.

The preliminary data on *Ectocarpus* sex chromosomes revealed substantial divergence from well studied plant or animal systems (Ahmed et al., unpublished). One of the interesting *Ectocarpus* features is that sex is determined during the haploid stage of the life cycle (UV chromosome types), in contrast to thoroughly studied systems with diploid sexes controlled by heteromorphic pairs of sex-chromosomes (ex. human (XY) or avian (ZW)).

Altogether, these unusual traits make *E. siliculosus* a very compelling model for future investigation of various evolutionary implications of haploid SDRs, regulatory network controlling sex determination and gamete size evolution, which we hope will be complemented by our data.

What have we learned from Ectocarpus transcriptome?

Our studies demonstate that the complexity of *Ectocarpus* gamete transcriptomes equal those of oogamous species (Masters *et al.*, 1992)(Wuest *et al.*, 2010)(Martins *et al.*, 2013) with remarkable divergence between males and females reflecting the early establishment of distinct sexual roles (Chapter 4). Over 4,000 genes were found with sex-biased expression pattern, enriched in microtubule based movement, vesicle trafficking, ion dynamics, cell wall biosynthesis, transcription and translation regulation, and signaling related categories. These high dimensional data are however composed of all expressed genes, for many of which there

is no corresponding biochemical activity or known function. Therefore, computer-intensive methods combined with wet-lab experiments are needed to understand the biological relevance of the information that has been generated.

Of special relevance are techniques with the capacity to "switch off" genes, either by genetic transformation or gene silencing. The latter should allow analyzing gene function in unprecedented detail in brown algae. As for today, and in the absence of such tools, our findings are based principally on sequence homology (Cock *et al.*, 2012). Nevertheless, our results established a first ground for testing various hypotheses about gamete molecular functioning and events that take place during interaction at fertilization, such as chemoreception, cell-cell recognition and fusion processes.

We highlighted interesting candidates for signaling genes that may be involved in gamete communication (REJ, GPCR, sulphonyltransferases and phosphatidylinositol signaling pathway related genes) as well as gamete biased transcription factors (MYB, NIN) that might be involved in sex specific gene expression regulation. However, the precise function of these genes will need to be studied in the future.

Evolution of sex-biased genes

The egg-sperm interactions at fertilization are usually subject to rapid evolutionary change (Vacquier, 1998) with the greatest variability being observed within gamete recognition proteins (Swanson & Vacquier, 2002)(Clark *et al.*, 2006). Structural diversity of fertilization genes implies their independent evolutionary origin, whereas genomic sequence variability between closely related species indicates involvement of adaptive evolutionary pressure (Vacquier, 1998).

Sea urchin sperm bindin, abalone sperm lysin protein, and vertebrate sperm fertilin are important components of the sperm-egg binding and fusion pathway, however there is no apparent homology between these three genes, suggesting they are autonomous products of evolution (Vacquier, 1998). Additionally, while some processes during gamete interaction seem to be conserved (like fusion of the acrosomal vesicle with the sperm plasma membrane), gamete recognition proteins are often under diversifying selection. Most pronounced examples include the marine invertebrate receptors bindin and lysin (Metz & Palumbi, 1996)(Yang *et al.*, 2000b).

Lack of sequence homology between gamete compatibility receptors makes it difficult to predict potential players at the recognition during fertilization when less-documented species are investigated. Therefore, signature of positive selection seems to be the best common denominator for this otherwise very diverse group of proteins.

Although, it was shown that the functional domains in the egg coat proteins of mollusk and yeast express the same 3D structure as human zona pellucida proteins, this relationship was established in the absence of sequence homology (Swanson *et al.*, 2011). The estimation of likelihood that a given protein matches a known 3D profile does not imply the common ancestry of compared proteins. Relationship between these three molecules could be a result of adaptation of a particular fold for the presentation of glycans that are involved in sperm-egg binding (Clark, 2013). More sequence data from related species will be needed to attribute the observed similarities to either direct homology or convergent evolution.

The widely occurring phenomenon of adaptive evolution in gamete receptor genes might have important consequences in the evolution of species (Swanson & Vacquier, 2002). Differences in amino acid sequence of bindin have been proposed to hinder cross-species hybridization in *Echinometra* (Metz & Palumbi, 1996). Such changes in receptor compatibility could be responsible for establishment of reproductive barriers and link gamete recognition genes to speciation (Palumbi, 2008).

Similar amino acid divergence was observed in candidate signaling receptors from *Ectocarpus* when geographically distant populations/species were investigated (Chapter 5). A strong signal of positive selection was clearly evident for Esi0130_0068, a protein with domain composition resembling sea urchin recognition for egg jelly (suREJ) receptor (Moy *et al.*, 1996b). suREJ triggers acrosomal reaction by binding the fucose sulfate polysaccharides on egg surface (Vacquier & Moy, 1997), a property that is particularly appealing, since lectin-like characteristics are expected from *Ectocarpus* gamete recognition receptor (Schmid, 1993).

In addition, three other male-biased genes were found with sites under episodic adaptive evolution (Esi0214_0046, Esi0050_0107 and Esi0146_0035), making our transcriptome data a good starting point for studies on directional selection and forces determining the level of diversity in the sex-related genes in *Ectocarpus*. Future projects using sequence information of full coding regions in a larger number of strains as well as experimental validation of gene function would provide deeper insight into adaptive changes and their consequences.

Algal lectins and biochemistry of gamete recognition

Fertilization in many marine species, including invertebrates and algae, occurs externally in the water column. Several strategies have been developed by these broadcast spawners to ensure successful gamete encounter, with chemoattraction playing one of the leading roles. In *Ectocarpus* species, settled female gametes release cyclo-olefinic hydrocarbons to attract swimming male gametes (Müller *et al.*, 1971)(Müller & Schmid, 1988). Pheromone response

is however, conserved within the genus, allowing for gametes of different species to allure each other despite established barriers to sexual fusion (Muller, 1979a).

These barriers might have arisen as a result of incompatibility of agglutinin-like sex-recognition proteins and their carbohydrate ligands (Schmid, 1993). To study the relation between the diversity of recognition proteins and their possible attribution to gamete incompatibility, we aimed at isolating gamete receptors from gametes of *Ectocarpus* New Zealand strain. However, no straightforward results were obtained.

Working with an *Ectocarpus* strain isolated from Port Aransas, used in the original study describing gamete receptor affinity (Schmid, 1993), could provide a more conclusive outcome in the future. Nonetheless, the gametophyte generation could not be recovered from the parthenosporophytes stock cultures for this particular strain. As for today, no gamete compatibility data for Port Aransas-New Zealand crosses are available. Therefore, it cannot be excluded that the recognition proteins of the two strains are already diverged enough to exhibit different carbohydrate affinities. Moreover, fertilization block assay using Port Aransas compatible lectin (Wheat Germ Agglutinin) did not affect gamete fusion in New Zealand strain, supporting this hypothesis.

No lectins have been described from brown algae to-date. Nevertheless, it has been noted that red and green algal lectins display strong specificity for extended carbohydrate structures and/or glycoproteins, often with no affinity for monosaccharides (Boyd *et al.*, 1997)(Hori *et al.*, 2000)(Bewley *et al.*, 2004)(Mori *et al.*, 2005)(Hori *et al.*, 2007). It might be possible then, that the isolation approach implemented in this study was not adequate for purification of a lectin with such specificity. Alternatively, the lectin-ligand interactions could have been powerful enough to inhibit elution from affinity columns using standard reagents (simple sugars, chaotropic ions, detergents and denaturants) as described by Hori et al. (2007).

Different properties of algal lectins comparing to plant or animal agglutinins, may also have had an effect on fertilization block assays. Much higher affinity of male gamete receptors to female exposed ligand could replace the plant lectins (used in the experiment to forbid fertilization) at the binding site, reversing the inhibition effect and consequently leading to fertilization within short time from gamete mixing point. Subsequent studies using lectins of broader affinity spectrum, including reported algal agglutinins, may help resolving that problem.

Most lectins from marine algae, are monomeric proteins with low molecular weight, comprising a new category of lectins (Hori *et al.*, 2000)(Ambrosio *et al.*, 2003)(Kim *et al.*, 2006) with promising properties in human health research. It was shown that binding of *Eucheuma serra* agglutinin induced apoptotic death in several human cancer cell lines (Sugahara *et al.*, 2001)(Fukuda *et al.*, 2006). Additionally, specificity for high-mannose N-

glycans of algal lectins makes them potential anti-viral agents, preventing infection with HIV (Bewley *et al.*, 1998)(Bewley *et al.*, 2004)(Mori *et al.*, 2005)(Botos & Wlodawer, 2005) or SARS coronavirus (O'Keefe *et al.*, 2010).
Concluding, algal lectins represent an interesting, largely unexplored group with compelling properties. However, further research is needed to uncover the involvement of these proteins in gamete coupling and other cellular processes in algae.

Technological progress has made today's biological studies more fascinating than at any time in the past. Genetic and genomic tools with Next Generation Sequencing and development in bioinformatics provide a jumping-off point for new discoveries, new concepts, and new wisdom. We now have the tools to study the evolution of sexual reproduction, sex determination, gamete recognition and speciation at a greater detail, to answer the endless list of questions, including those raised by discoveries yet to come.

Summary

Despite its prevalence among eukaryotes, there are still many questions awaiting answers and plenty of discoveries to be made with regard to the mechanisms and maintenance of sexual reproduction. Although the molecular basis of sex has been studied in great detail in a limited number of model organisms, it remains to be determined whether the mechanisms described in model systems are conserved in other species.

Brown algae (Phaeophyceae) represent a very interesting group to study the evolution of sexual reproduction, owing to the rich variation of life cycles, fertilization modes and sex determination systems displayed by this lineage. Particularly unique are various degrees of sexual dimorphism between gametes of closely related species, ranging from isogamy through anisogamy to oogamy. It is generally accepted that the evolution of morphologically identical isogametes has preceded the evolution of dimorphic sperm and egg cells, which led to nearly universal domination of oogamy in multicellular organisms. Therefore, the opportunity to compare these systems on the molecular level in Phaeophyceae promises advancement in our understanding of the evolution of sexual reproduction not only in algae, but in all living organisms.

The technological progress in genomic research presented us with new possibilities in studying large-scale gene expression, uncovering the secrets of how genomes work and unlocking basic biological functions. Today, we are able to identify thousands of genes that are associated with a specific biological activity, process or change. These advanced techniques are now being put into use in algal research, starting with the establishment of the whole genome sequence of *Ectocarpus siliculosus*.

Ectocarpus is a brown, isogamous, filamentous alga inhabiting temperate, coastal environments around the world. Many aspects of *E. siliculosus* life cycle (including fertilization, early development and transition of generations) have been researched in the past, providing a valuable foundation for evolution of sexual reproduction studies.

In this thesis we investigated the molecular basis of sexual reproduction and its link to reproductive isolation in brown algae using a New Zealand strain of *Ectocarpus* starting with characterization of gamete recognition receptors (Chapter 2). Based on the described carbohydrate and lectin affinities of *E. siliculosus* cell-cell agglutinins, we performed chromatography experiments aimed at isolation of these proteins from gamete extracts. Since reproductive proteins have been shown to undergo rapid diversification, they are thought to be crucial factors in establishment of reproductive barriers and speciation. Thus,

combining sequence comparison and knowledge of sexual compatibility could provide a valuable insight into the link between reproductive isolation and gamete recognition in *Ectocarpus*. Unfortunately, no convincing recognition proteins could be recovered, leading to employment of an alternative, genomic approach.

Thereto, we constructed whole transcriptome libraries of male and female gametes to explore their functional differentiation at the gene expression level (Chapter 3 and Chapter 4). Next Generation Sequencing technology (SOLiD) was used to quantitatively assign transcribed genes to each gamete type. A procedure of short-read data analysis and its evaluation is described in Chapter 3. This extensive and powerful technique allowed for identification of more than 4,000 genes with differential expression pattern between male and female gametes within many important cellular processes, including pheromone production, cell movement, cell wall biosynthesis and others (Chapter 4). Moreover, signaling pathways related ontologies were well represented among gamete specific genes, allowing for selection of potential cell-cell recognition candidates. These genes were then sequenced in a representative number of strains from distant geographical locations and varying stages of reproductive isolation, to search for signatures of positive selection.

Maximum likelihood analyses using PAML and HyPhy classified numerous amino acid sites as subjected to adaptive evolution in at least four of the studied male-biased genes (Chapter 5). Observed variation may be an indication of direct involvement of these genes in gamete interaction during reproduction. Interestingly, one of the identified genes (Esi0130_0068) was similar in topology to the egg recognition protein in sea urchin sperm, making it an attractive candidate for future experimental evaluation.

Taken together, this thesis has offered the first insight into protist gametes transcriptomes, revealing complex, functional organization and differentiation according to the adopted sexual roles, as well as evolutionary variation of sex biased genes. Our study supports the relevance of algal systems in the research of sexual reproduction and provides a starting point for more exciting discoveries in that field.

Samenvatting

Ondanks het feit dat nagenoeg alle eukaryoten zich seksueel voortplanten, zijn er nog talloze open vragen met betrekking tot de mechanismen en het behoud van seksuele reproductie. De moleculaire basis van seksuele reproductie werd weliswaar in detail bestudeerd in een beperkt aan modelorganismen, maar het is verre van duidelijk tot op welk niveau de beschreven mechanismen ook geconserveerd zijn in andere soorten.

Omwille van hun grote variatie in levenscycli, vormen van fertilisatie en seksdeterminatiesystemen vormen Bruinwieren (Phaeophyceae) een bijzonder interessante groep voor de studie van seksuele reproductie. Uniek binnen de eukrayoten zijn de diverse gradaties van seksueel dimorfisme in gameten tussen nauw verwante soorten, variërend van isogamie over anisogamie tot oogamie, die men aantreft bij bruinwieren. Er wordt algemeen aanvaard dat seksueel dimorphisme in gameten, geëvolueerd is uit isogamie, wat uiteindelijk resulteerde in een dominantie van oogamie, eicellen en spermacellen, in meercellige eukaryoten. De mogelijkheid om deze systemen op moleculair niveau te vergelijken binnen de Phaeophyceae belooft niet enkel vooruitgang in onze kennis van de evolutie van seksuele reproductie bij algen, maar in alle levende organismen. Technologische vooruitgang in genomica biedt nieuwe mogelijkheden om genexpressie te bestuderen. Door middel van genexpressiestudies kan een beter inzicht verkregen worden in de werking van het genoom en de elementaire biologische functies. De huidige technologie laat toe om relatief snel de identiteit van duizenden genen te bepalen, geassocieerd met specifieke biologische activiteiten, processen en veranderingen. De beschikbaarheid van volledige genoomsequenties, bijvoorbeeld van Ectocarpus siliculosus, zorgt ervoor dat dergelijke innovatieve technieken ook momenteel geïmplementeerd kunnen worden algologisch onderzoek.

Ectocarpus is een isogaam, filamentous bruinwier dat algemeen voorkomt langsheen kusten in gematigde regios. Een hele reeks aspecten met betrekking tot de levenscyclus van Ectocarpus (b.v. fertilisatie vroege ontwikkeling en generatiewisseling) werden reeds in detail bestudeerd in het verleden en vormen een solide basis voor verdere studies naar de evolutie van seksuele voortplanting. In deze scriptie bestudeer ik de moleculaire basis van seksuele voortplanting en daarmee gepaard gaande reproductieve isolatie in bruinwieren aan de hand van een Ectocarpus strain uit Nieuw-Zeeland. Er wordt eerst getracht de gameetherkenningsreceptoren te karakteriseren (Hoofdstuk 2) aan de hand van de

carbohydraat- en lectine-affiniteiten bij Ectocarpus. Met behulp van chromatografie experimenten werd gepoogd de gameetherkenningseiwitten te isoleren uit gameetextracten. Aangezien gameetherkenningseiwitten zeer snel evolueren worden ze geacht een cruciale functie te hebben bij het tot stand komen van reproductieve isolatie en speciatie. Een vergelijk tussen sequentiedivergentie en seksuele compatibiliteit kan daarom belangrijke inzichten leveren met betrekking tot reproductieve isolatie en gameetherkenning bij Ectocarpus. Helaas bracht deze piste geen overtuigende gameetherkenningseiwitten aan het licht, wat dan weer aanleiding gaf tot een genomische benadering van deze onderzoeksvraag. Transcriptoom-libraries werden geconstrueerd van mannelijke en vrouwelijke gameten om hun functionele differentiatie te bestuderen aan de hand van sex-specifieke genexpressiepatronen (Hoofdstuk 3 en 4). Aan de hand van Next Generation Sequencing technology (SOLiD) werd de genexpressie in elk gameettype quantitatief bepaald. De analyse van korte SOLiD reads werd beschreven in Hoofdstuk 3. Deze krachtige techniek maakt het mogelijk of differentiële expressie van meer dan 4000 genen aan te tonen tussen mannelijke en vrouwelijke gameten. Ook op het niveau van cellulaire processen, bijvoorbeeld feromoonproductie, motiliteit en celwand biosynthese, werden belangrijke verschillen aangetoond (Hoofdstuk 4). Genen met aangetoonde functies in signalling pathways waren eveneens goed vertegenwoordigd, wat toeliet om potentiële cel-cel herkenningseiwitten te identificeren. Deze genen werden vervolgens geamplificeerd en gesequeneerd in een representatief aantal Ectocarpus strains van verschillende locaties en met verschillende gradaties van reproductieve isolatie, om aanwijzingen te vinden voor positieve selectie op gameetherkenningseiwitten. Maximum likelihood analyses met PAML en HyPhy wezen verschillende aminozuurposities aan in minstens vier van de bestudeerde genen die overgeëxpresseerd werden in mannelijke gameten (Hoofdstuk 5). Deze resultaten wijzen mogelijk op een directe betrokkenheid van deze genen in gameetinteracties tijdens seksuele reproductie. Een van de geïdentificeerde genen (Esi0130_0068) heeft een gelijkaardige structuur als een goed bestudeerd gameetherkenningseiwit in zeeëgels en wordt voorgesteld als een interessant kandidaateiwit voor toekomstige experimentele studies.

In conclusie, deze scriptie vormt de eerste transcriptoomstudie van gameten bij protisten. De resultaten wijzen op een complexe functionele organisatie en differentiatie van mannelijke en vrouwelijke gameten. Aanvullend wijst de hoge differentiatie in genexpressie tussen mannelijke en vrouwelijke isogameten op het feit dat differentiatie op cellulair niveau vooruitloopt op morfologische differentiatie. De resultaten tonen verder de relevantie aan van algen voor onderzoek naar seksuele reproductie bij eukaryoten en vormen een aanlokkelijk startpunt voor verder onderzoek.

Bibliography

Ainsworth C. **2000**. Boys and girls come out to play: the molecular biology of dioecious plants. *Annals of Botany* **86**: 211–221.

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. **1997**. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research* **25**: 3389–3402.

Ambrosio AL, Sanz L, Sánchez EI, Wolfenstein-Todel C, Calvete JJ. **2003**. Isolation of two novel mannan- and l-fucose-binding lectins from the green alga Enteromorpha prolifera: biochemical characterization of EPL-2. *Archives of Biochemistry and Biophysics* **415**: 245–250.

Ameur A, Zaghlool A, Halvardson J, Wetterbom A, Gyllensten U, Cavelier L, Feuk L. **2011**. Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nature Structural & Molecular Biology* **18**: 1435–1440.

Araki Y, Shimizu HD, Saeki K, Okamoto M, Yamada L, Ishida K, Sawada H, Urushihara H. **2012**. A surface glycoprotein indispensable for gamete fusion in the social amoeba Dictyostelium discoideum. *Eukaryotic Cell* **11**: 638–644.

Archibald JM. **2009**. The puzzle of plastid evolution. *Current Biology* **19**: R81–R88.

Archibald JM. **2012**. The evolution of algae by secondary and tertiary endosymbiosis. Advances in Botanical Research. Elsevier, 87–118.

Armbrust EV. **1999**. Identification of a new gene family expressed during the onset of sexual reproduction in the centric diatom Thalassiosira weissflogii. *Applied and Environmental Microbiology* **65**: 3121–3128.

Armbrust EV, Galindo HM. **2001**. Rapid evolution of a sexual reproduction gene in centric diatoms of the genus Thalassiosira. *Applied and Environmental Microbiology* **67**: 3501–3513.

Arun A, Peters NT, Scornet D, Peters AF, Mark Cock J, Coelho SM. **2013**. Non-cell autonomous regulation of life cycle transitions in the model brown alga Ectocarpus. *New Phytologist* **197**: 503–510.

Assis R, Zhou Q, Bachtrog D. **2012**. Sex-biased transcriptome evolution in drosophila. *Genome Biology and Evolution* **4**: 1189–1200.

Avise JC, Nicholson TH. **2011**. *Hermaphroditism: a primer on the biology, ecology, and evolution of dual sexuality*. Columbia University Press.

Bachtrog D, Kirkpatrick M, Mank JE, McDaniel SF, Pires JC, Rice W, Valenzuela N. **2011**. Are all sex chromosomes created equal? *Trends in Genetics* **27**: 350–357.

Le Bail A, Billoud B, Kowalczyk N, Kowalczyk M, Gicquel M, Le Panse S, Stewart S, Scornet D, Cock JM, Ljung K, *et al.* Auxin metabolism and function in the multicellular brown alga Ectocarpus siliculosus. *Plant Physiol.* **153**: 128–144.

Le Bail A, Billoud B, Maisonneuve C, Peters AF, Cock JM, Charrier B. **2008**. Early development pattern of the brown alga Ectocarpus siliculosus (Ectocarpales, Phaeophyceae) sporophyte. *J. Phycol.* **44**: 1269–1281.

Baker JRJ, Evans LV. **1973a**. The ship fouling alga Ectocarpus. I. Ultrastructure and Cytochemistry of Plurilocular Reproductive Stages. *Protoplasma* **77**: 1–13.

Baker DJRJ, Evans LV. **1973b**. The ship-fouling algaEctocarpus. II. Ultrastructure of the unilocular reproductive stages. *Protoplasma* **77**: 181–189.

Baldauf SL. **2003**. The deep roots of Eukaryotes. *Science* **300**: 1703–1706.

Barash D. **1976**. What does sex really cost? *American Naturalist* **110**: 894–897.

Barton NH. **1995**. A general model for the evolution of recombination. *Genetical Research* **65**: 123–144.

Barton NH, Briggs DEG, Eisen JA, Goldstein DB, Patel NH. **2007**. *Evolution*. Cold Spring Harbor Laboratory Press.

Barton NH, Charlesworth B. **1998**. Why sex and recombination? *Science* **281**: 1986–1990.

Basile G, Aker M, Mortimer RK. **1992**. Nucleotide sequence and transcriptional regulation of the yeast recombinational repair gene RAD51. *Molecular and cellular biology* **12**: 3235–3246.

Basu S, Sun H, Brian L, Quatrano RL, Muday GK. **2002**. Early embryo development in Fucus distichus is auxin sensitive. *Plant Physiology* **130**: 292–302.

Beck CF, Acker A. **1992**. Gametic differentiation of Chlamydomonas reinhardtii : control by nitrogen and light. *PLANT PHYSIOLOGY* **98**: 822–826.

Beck CF, Haring MA. **1996**. Gametic differentiation of Chlamydomonas. International Review of Cytology. Elsevier, 259–302.

Bell G. **1978**. The evolution of anisogamy. *Journal of Theoretical Biology* **73**: 247–270.

Bell G. **1982**. *The masterpiece of nature: the evolution and genetics of sexuality*. Univ of California Pr.

Bell G. **1997**. The evolution of the life cycle of brown seaweeds. *Biological Journal of the Linnean Society* **60**: 21–38.

Bernstein H, Byerly HC, Hopf FA, Michod RE. **1985**. Genetic damage, mutation, and the evolution of sex. *Science (New York, N.Y.)* **229**: 1277–1281.

Bernstein H, Hopf F, Michod R. **1987**. The molecular basis of the evolution of sex. Advances in Genetics. Elsevier, 323–370.

Bernstein H, Hopf F, Michod R. **1988**. Is meiotic recombination an adaptation for repairing DNA, producing genetic variation, or both? The Evolution of sex: an examination of current ideas, edited by R. Michod and B. Levin. Sinauer Associates, 194–211.

Bewley CA, Cai M, Ray S, Ghirlando R, Yamaguchi M, Muramoto K. **2004**. New carbohydrate specificity and HIV-1 fusion blocking activity of the cyanobacterial protein MVL: NMR, ITC and sedimentation equilibrium studies. *Journal of Molecular Biology* **339**: 901–914.

Bewley CA, Gustafson KR, Boyd MR, Covell DG, Bax A, Clore GM, Gronenborn AM. **1998**. Solution structure of cyanovirin-N, a potent HIV-inactivating protein. *Nature structural biology* **5**: 571–578.

Birky CW. **2001**. The inheritance of genes in mitochondria and chloroplasts: Laws, mechanisms and models. *Annual Review of Genetics* **35**: 125–148.

Blankenberg D, Gordon A, Von Kuster G, Coraor N, Taylor J, Nekrutenko A, Galaxy Team. **2010**. Manipulation of FASTQ data with Galaxy. *Bioinformatics (Oxford, England)* **26**: 1783–1785.

Blankenberg D, Kuster GV, Coraor N, Ananda G, Lazarus R, Mangan M, Nekrutenko A, Taylor J. **2001**. Galaxy: a web-based genome analysis tool for experimentalists. Current Protocols in Molecular Biology. John Wiley & Sons, Inc., .

Blier P. **2001**. Natural selection and the evolution of mtDNA-encoded peptides: evidence for intergenomic co-adaptation. *Trends in Genetics* **17**: 400–406.

Blouin NA, Brodie JA, Grossman AC, Xu P, Brawley SH. **2011**. Porphyra: a marine crop shaped by stress. *Trends in Plant Science* **16**: 29–37.

Boland W. **1995**. The chemistry of gamete attraction: chemical structures, biosynthesis, and (a)biotic degradation of algal pheromones. *Proceedings of the National Academy of Sciences* **92**: 37–43.

Boland W, Marner F-J, Jaenicke L, Muller DG, Folster E. **1983**. Comparative receptor study in gamete chemotaxis of the seaweeds Ectocarpus siliculosus and Cutleria multifida. An approach to interspecific communication of algal gametes. *European Journal of Biochemistry* **134**: 97–103.

Boland W, Mertes K. **1985**. Biosynthesis of algal pheromones. *European Journal of Biochemistry* **147**: 83–91.

Bolwell GP, Callow JA, Callow ME, Evans LV. **1979**. Fertilization in brown algae. II. Evidence for lectin-sensitive complementary receptors involved in gamete recognition in Fucus serratus. *Journal of Cell Science* **36**: 19–30.

Bothwell JH, Marie D, Peters AF, Cock JM, Coelho SM. **2010**. Role of endoreduplication and apomeiosis during parthenogenetic reproduction in the model brown alga Ectocarpus. *New Phytologist* **188**: 111–121.

Botos I, Wlodawer A. **2005**. Proteins that bind high-mannose sugars of the HIV envelope. *Progress in biophysics and molecular biology* **88**: 233–282.

Bowler C, Allen AE, Badger JH, Grimwood J, Jabbari K, Kuo A, Maheswari U, Martens C, Maumus F, Otillar RP, *et al.* 2008. The Phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature* **456**: 239–244.

Boyd MR, Gustafson KR, McMahon JB, Shoemaker RH, O'Keefe BR, Mori T, Gulakowski RJ, Wu L, Rivera MI, Laurencot CM, *et al.* 1997. Discovery of cyanovirin-N, a novel human immunodeficiency virus-inactivating protein that binds viral surface envelope glycoprotein gp120: potential applications to microbicide development. *Antimicrobial Agents and Chemotherapy* **41**: 1521–1530.

Bradford MM. **1976**. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry* **72**: 248–254.

Bram MR. **2002**. Effects of inbreeding in three populations of the dioecious annual Amaranthus cannabinus (Amaranthaceae). *Journal of the Torrey Botanical Society* **129**: 298.

Brawley S, Johnson L, Pearson G, Speransky V, Li R, Serrao E. **1999**. Gamete release at low tide in fucoid algae: Maladaptive or advantageous? *American Zoologist* **39**: 218–229.

Brownlee C, Bouget F-Y, Corellou F. **2001**. Choosing sides: establishment of polarity in zygotes of fucoid algae. *Seminars in Cell & Developmental Biology* **12**: 345–351.

Bull JJ. **1983**. *Evolution of sex determining mechanisms*. Benjamin/Cummings Pub. Co.

Bustamante CD, Fledel-Alon A, Williamson S, Nielsen R, Hubisz MT, Glanowski S, Tanenbaum DM, White TJ, Sninsky JJ, Hernandez RD, *et al.* 2005. Natural selection on protein-coding genes in the human genome. *Nature* **437**: 1153–1157.

Carvalho AB. **2002**. Origin and evolution of the Drosophila Y chromosome. *Current Opinion in Genetics & Development* **12**: 664–668.

Cavalier-Smith T. **1980**. How selfish is DNA? *Nature* **285**: 617–618.

Cavalier-Smith T. **2002**. Origins of the machinery of recombination and sex. *Heredity* **88**: 125.

Chaisson MJ, Brinza D, Pevzner PA. **2009**. De novo fragment assembly with short mate-paired reads: Does the read length matter? *Genome Research* **19**: 336–346.

Chapman T, Liddle LF, Kalb JM, Wolfner MF, Partridge L. **1995**. Cost of mating in Drosophila melanogaster females is mediated by male accessory gland products. *Nature* **373**: 241–244.

Charlesworth B. **1978**. The population genetics of anisogamy. *Journal of theoretical biology* **73**: 347–357.

Charlesworth B. **1980**. The cost of sex in relation to mating system. *Journal of Theoretical Biology* **84**: 655–671.

Charlesworth B. **1990**. Mutation-selection balance and the evolutionary advantage of sex and recombination. *Genetical Research* **55**: 199–221.

Charlesworth B. **1996**. The evolution of chromosomal sex determination and dosage compensation. *Current Biology* **6**: 149–162.

Charlesworth D. **2013**. Plant sex chromosome evolution. *Journal of Experimental Botany* **64**: 405–420.

Charlesworth B, Charlesworth D. **1978**. A model for the evolution of dioecy and gynodioecy. *The American Naturalist* **112**: 975.

Charlesworth D, Charlesworth B. **1980**. Sex differences in fitness and selection for centric fusions between sex-chromosomes and autosomes. *Genetics Research* **35**: 205–214.

Charlesworth B, Charlesworth D. **1999**. The genetic basis of inbreeding depression. *Genetical Research* **74**: 329–340.

Charlesworth D, Charlesworth B. **2005**. Sex chromosomes: evolution of the weird and wonderful. *Current Biology* **15**: R129–R131.

Charlesworth D, Charlesworth B, Marais G. **2005**. Steps in the evolution of heteromorphic sex chromosomes. *Heredity* **95**: 118–128.

Charlesworth D, Morgan MT. **1991**. Allocation of resources to sex functions in flowering plants. *Philosophical Transactions of the Royal Society B: Biological Sciences* **332**: 91–102.

Charrier B, Coelho SM, Le Bail A, Tonon T, Michel G, Potin P, Kloareg B, Boyen C, Peters AF, Cock JM. **2008**. Development and physiology of the brown alga Ectocarpus siliculosus: two centuries of research. *New Phytologist* **177**: 319–332.

Chasnov JR. **2000**. Mutation-selection balance, dominance and the maintenance of sex. *Genetics* **156**: 1419–1425.

Chepurnov VA, Mann DG, Von Dassow P, Vanormelingen P, Gillard J, Inzé D, Sabbe K, Vyverman W. **2008**. In search of new tractable diatoms for experimental biology. *BioEssays* **30**: 692–702.

Chepurnov VA, Mann DG, Sabbe K, Vyverman W. **2004**. Experimental studies on sexual reproduction in diatoms. International Review of Cytology. Elsevier, 91–154.

Civetta A, Singh RS. **1998**. Sex-related genes, directional sexual selection, and speciation. *Molecular Biology and Evolution* **15**: 901–909.

Clark G. **1981**. *Staining procedures*. Published for the Biological Stain Commission by Williams & Wilkins.

Clark GF. **2013**. The role of carbohydrate recognition during human sperm–egg binding. *Human Reproduction* **28**: 566–577.

Clark NL, Aagaard JE, Swanson WJ. **2006**. Evolution of reproductive proteins from animals and plants. *Reproduction* **131**: 11–22.

Clark AG, Begun DJ, Prout T. **1999**. Female × male interactions in Drosophila sperm competition. *Science* **283**: 217–220.

Clark NL, Gasper J, Sekino M, Springer SA, Aquadro CF, Swanson WJ. **2009**. Coevolution of interacting fertilization proteins. *Plos Genetics* **5**.

Cleveland LR. **1947**. The origin and evolution of meiosis. *Science* **105**: 287–289.

Cock JM, Peters AF, Coelho SM. **2011**. Brown algae. *Current Biology* **21**: R573–R575.

Cock JM, Sterck L, Ahmed S, Allen AE, Amoutzias G, Anthouard V, Artiguenave F, Arun A, Aury J-M, Badger JH, *et al.* 2012. The Ectocarpus genome and brown algal genomics. Advances in Botanical Research. Elsevier, 141–184.

Cock JM, Sterck L, Rouze P, Scornet D, Allen AE, Amoutzias G, Anthouard V, Artiguenave F, Aury JM, Badger JH, *et al.* 2010. The Ectocarpus genome and the independent evolution of multicellularity in brown algae. *Nature* **465**: 617–621.

Coelho SM, Godfroy O, Arun A, Le Corguillé G, Peters AF, Cock JM. **2011**. OUROBOROS is a master regulator of the gametophyte to sporophyte life cycle transition in the brown alga Ectocarpus. *Proceedings of the National Academy of Sciences*.

Coelho SM, Scornet D, Rousvoal S, Peters NT, Dartevelle L, Peters AF, Cock JM. **2012**. How to cultivate Ectocarpus. *Cold Spring Harbor Protocols* **2012**: pdb.prot067934.

Coelho SM, Simon N, Ahmed S, Cock JM, Partensky F. **2013**. Ecological and evolutionary genomics of marine photosynthetic organisms. *Molecular Ecology* **22**: 867–907.

Colegrave N. **2002**. Sex releases the speed limit on evolution. *Nature* **420**: 664–666.

Colegrave N. **2012**. The evolutionary success of sex. *EMBO reports* **13**: 774–778.

Conesa A, Gotz S, Garcia-Gomez JM, Terol J, Talon M, Robles M. **2005**. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**: 3674–3676.

Cosmides L, Tooby J. **1981**. Cytoplasmic inheritance and intragenomic conflict. *Journal of Theoretical Biology* **89**: 83–129.

Costich DE. **1995**. Gender specialization across a climatic gradient: experimental comparison of monoecious and dioecious Ecballium. *Ecology* **76**: 1036.

Costich DE, Meagher TR. **1992**. Genetic variation in Ecballium elaterium (Cucurbitaceae): Breeding system and geographic distribution. *Journal of Evolutionary Biology* **5**: 589–601.

Cox MM. **1993**. Relating biochemistry to biology: how the recombinational repair function of RecA protein is manifested in its molecular properties. *BioEssays: news and reviews in molecular, cellular and developmental biology* **15**: 617–623.

Coyne JA, Orr HA. **1989**. Patterns of speciation in Drosophila. *Evolution* **43**: 362.

Coyne JA, Orr HA. **2004**. *Speciation*. New York: Sinauer.

Crow JF. **1988**. The importance of recombination. The evolution of sex: An examination of current ideas (Michod, R.E. and Levin, B.R., eds). Sinauer Associates, 56–73.

Dacks J, Roger AJ. **1999**. The first sexual lineage and the relevance of facultative sex. *Journal of Molecular Evolution* **48**: 779–783.

Darwin C. **1859**. *On the origin of species by means of natural selection, or, the preservation of favoured races in the struggle for life*.

Darwin C. **1872**. *The descent of man, and selection in relation to sex*. D. Appleton.

Dawkins R. **1989**. *The selfish gene*. Oxford University Press.

Delph LF. **2003**. Sexual dimorphism in gender plasticity and its consequences for breeding system evolution. *Evolution and Development* **5**: 34–39.

Delport W, Poon AFY, Frost SDW, Kosakovsky Pond SL. **2010**. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics (Oxford, England)* **26**: 2455–2457.

Destombe C, Cembella A. **1990**. Mating-type determination, gametic recognition and reproductive success in Alexandrium excavatum (Gonyaulacales, Dinophyta), a toxic red-tide dinoflagellate. *Phycologia* **29**: 316–325.

Dittami S. Ectocarpus transcriptomics homepage (http://www.sb-roscoff.fr/UMR7139/ectocarpus/transcriptomics/index.html).

Dittami SM, Scornet D, Petit JL, Segurens B, Da Silva C, Corre E, Dondrup M, Glatting KH, Konig R, Sterck L, *et al.* 2009. Global expression analysis of the brown alga Ectocarpus siliculosus (Phaeophyceae) reveals large-scale reprogramming of the transcriptome in response to abiotic stress. *Genome Biology* **10**: R66.

Dobzhansky T. **1936**. Studies on hybrid sterility. II. Localization of sterility factors in Drosophila pseudoobscura hybrids. *Genetics* **21**: 113–135.

Dobzhansky TG. **1937**. *Genetics and the origin of species.* Columbia University Press.

Dobzhansky T. **1951**. *Genetics and the origin of species. 3rd ed.* New York: Columbia Univ. Press.

Dorken ME, Friedman J, Barrett SCH. **2002**. The evolution and maintenance of monoecy and dioecy in Sagittaria latifolia (Alismataceae). *Evolution* **56**: 31–41.

Dougherty EC. **1955**. Comparative evolution and the origin of sexuality. *Systematic Zoology* **4**: 145.

Dowling D, Friberg U, Lindell J. **2008**. Evolutionary implications of non-neutral mitochondrial genetic variation. *Trends in Ecology & Evolution* **23**: 546–554.

Dusenbery DB. **2000**. Selection for high gamete encounter rates explains the success of male and female mating types. *Journal of Theoretical Biology* **202**: 1–10.

Eady PE, Hamilton L, Lyons RE. **2007**. Copulation, genital damage and early death in Callosobruchus maculatus. *Proceedings of the Royal Society B: Biological Sciences* **274**: 247–252.

Eberhard WG. **1996**. *Female control: sexual selection by cryptic female choice (monographs in behavior and ecology)*. Princeton University Press.

Edlund MB, Stoermer EF. **1997**. Ecological, evolutionary, and systematic significance of diatom life histories. *Journal of Phycology* **33**: 897–918.

Elder J, Turner B. **1995**. Concerted evolution of repetitive DNA sequences in eukaryotes. *Quarterly Review of Biology* **70**: 297–320.

Elias M, Archibald JM. **2009**. Sizing up the genomic footprint of endosymbiosis. *BioEssays* **31**: 1273–1279.

Engelstädter J. **2008**. Constraints on the evolution of asexual reproduction. *BioEssays* **30**: 1138–1150.

Eppley SM, Jesson LK. **2008**. Moving to mate: the evolution of separate and combined sexes in multicellular organisms. *Journal of Evolutionary Biology* **21**: 727–736.

Eppley SM, Pannell JR. **2007**. Density-dependent self-fertilization and male versus hermaphrodite siring success in an androdioecious plant. *Evolution* **61**: 2349–2359.

Esser C. **2004**. A genome phylogeny for mitochondria among -proteobacteria and a predominantly eubacterial ancestry of yeast nuclear genes. *Molecular Biology and Evolution* **21**: 1643–1660.

Ewing B, Hillier L, Wendl MC, Green P. **1998**. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome research* **8**: 175–185.

Fallis LC, Stein KK, Lynn JW, Misamore MJ. **2010**. Identification and role of carbohydrates on the durface of gametes in the zebra mussel, Dreissena polymorpha. *Biol. Bull.* **218**: 61–74.

Felsenstein J. **1981**. Evolutionary trees from DNA sequences: A maximum likelihood approach. *Journal of Molecular Evolution* **17**: 368–376.

Ferris PJ, Goodenough UW. **1997**. Mating type in Chlamydomonas is specified by mid, the minus-dominance gene. *Genetics* **146**: 859–869.

Ferris P, Olson B, De Hoff PL, Douglass S, Casero D, Prochnik S, Geng S, Rai R, Grimwood J, Schmutz J, *et al.* 2010. Evolution of an expanded sex-determining locus in Volvox. *Science* **328**: 351–354.

Ferris PJ, Pavlovic C, Fabry S, Goodenough UW. **1997**. Rapid evolution of sex-related genes in Chlamydomonas. *Proceedings of the National Academy of Sciences of the United States of America* **94**: 8634–8639.

Fiers W, Contreras R, Duerinck F, Haegeman G, Iserentant D, Merregaert J, Min Jou W, Molemans F, Raeymaekers A, Van den Berghe A, *et al.* 1976. Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. *Nature* **260**: 500–507.

Fisher RA. **1930**. *The genetical theory of natural selection: a complete variorum edition*. Oxford University Press.

Fraser JA, Heitman J. **2005**. Chromosomal sex-determining regions in animals, plants and fungi. *Current Opinion in Genetics & Development* **15**: 645–651.

Fu G, Nagasato C, Ito T, Müller DG, Motomura T. **2013**. Ultrastructural analysis of flagellar development in plurilocular sporangia of Ectocarpus siliculosus (Phaeophyceae). *Protoplasma* **250**: 261–272.

Fukuda Y, Sugahara T, Ueno M, Fukuta Y, Ochi Y, Akiyama K, Miyazaki T, Masuda S, Kawakubo A, Kato K. **2006**. The anti-tumor effect of Euchema serra agglutinin on colon cancer cells in vitro and in vivo: *Anti-Cancer Drugs* **17**: 943–947.

Geller A, Mueller DG. **1981**. Analysis of the flagellar beat pattern of male Ectocarpus siliculosus gametes (Pheophyta) in relation to chemotactic stimulation by female cells. *j. exp. Biol.* **92**: 53–66.

Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J, *et al.* 2005. Galaxy: A platform for interactive large-scale genome analysis. *Genome Research* **15**: 1451–1455.

Gillard J. **2009**. The pennate diatom life cycle: a generic, physiological and biochemical study using Seminavis robusta as a new experimental model. PhD thesis, Ghent University, Ghent, Belgium.

Gillham NW, Boynton JE, Johnson AM, Burkhart BD. **1987**. Mating type linked mutations which disrupt the uniparental transmission of chloroplast genes in Chlamydomonas. *Genetics* **115**: 677–684.

Glabe C. **1979**. Species-specific sperm adhesion in sea urchins. A quantitative investigation of bindin- mediated egg agglutination. *The Journal of Cell Biology* **83**: 595–604.

Gleiser G, Verdú M. **2004**. Repeated evolution of dioecy from androdioecy in Acer. *New Phytologist* **165**: 633–640.

Goddard MR, Godfray HCJ, Burt A. **2005**. Sex increases the efficacy of natural selection in experimental yeast populations. *Nature* **434**: 636–640.

Goecks J, Nekrutenko A, Taylor J, The Galaxy Team. **2010**. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology* **11**: R86.

Goldman N, Yang Z. **1994**. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Molecular biology and evolution* **11**: 725–736.

Goodenough UW. **1995**. Molecular genetics of sexuality in Chlamydomonas. *Annual Review of Plant Physiology and Plant Molecular Biology* **46**: 21–44.

Goodenough U, Lin H, Lee J-H. **2007**. Sex determination in Chlamydomonas. *Seminars in Cell & Developmental Biology* **18**: 350–361.

Graham MH, Kinlan BP, Druehl LD, Garske LE, Banks S. **2007**. Deep-water kelp refugia as potential hotspots of tropical marine diversity and productivity. *Proceedings of the National Academy of Sciences* **104**: 16576–16580.

Gray MW. **1999**. Mitochondrial evolution. *Science* **283**: 1476–1481.

Grun P. **1976**. *Cytoplasmic genetics and evolution*. Columbia Univ Pr.

Gunaratne HJ, Moy GW, Kinukawa M, Miyata S, Mah SA, Vacquier VD. **2007**. The 10 sea urchin receptor for egg jelly proteins (SpREJ) are members of the polycystic kidney disease-1 (PKD1) family. *BMC Genomics* **8**: 235.

Hadjivasiliou Z, Pomiankowski A, Seymour RM, Lane N. **2012**. Selection for mitonuclear co-adaptation could favour the evolution of two sexes. *Proceedings of the Royal Society B: Biological Sciences* **279**: 1865–1872.

Hamilton WD, Axelrod R, Tanese R. **1990**. Sexual reproduction as an adaptation to resist parasites (a review). *Proceedings of the National Academy of Sciences* **87**: 3566–3573.

Hamm D, Mautz BS, Wolfner MF, Aquadro CF, Swanson WJ. **2007**. Evidence of amino acid diversity-enhancing selection within humans and among primates at the candidate sperm-receptor gene PKDREJ. *American Journal of Human Genetics* **81**: 44–52.

Han JW, Klochkova TA, Shim JB, Yoon K, Kim GH. **2012**. Isolation and characterization of a sex-specific lectin in a marine red alga Aglaothamnion oosumiense Itono. *Applied and Environmental Microbiology* **78**: 7283–7289.

Heesch S, Cho GY, Peters AF, Le Corguillé G, Falentin C, Boutet G, Coëdel S, Jubin C, Samson G, Corre E, *et al.* 2010. A sequence-tagged genetic map for the brown alga Ectocarpus siliculosus provides large-scale assembly of the genome sequence. *New Phytologist* **188**: 42–51.

Hellberg ME, Vacquier VD. **1999**. Rapid evolution of fertilization selectivity and lysin cDNA sequences in teguline gastropods. *Molecular Biology and Evolution* **16**: 839–848.

Hellemans J, Mortier G, Paepe AD, Speleman F, Vandesompele J. **2007**. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biology* **8**: R19.

Hirai M, Arai M, Mori T, Miyagishima S-Y, Kawai S, Kita K, Kuroiwa T, Terenius O, Matsuoka H. **2008**. Male fertility of malaria parasites is determined by GCS1, a plant-type reproduction factor. *Current biology: CB* **18**: 607–613.

Hoekstra RF. **1982**. On the asymmetry of sex: Evolution of mating types in isogamous populations. *Journal of Theoretical Biology* **98**: 427–451.

Hoekstra RF. **1987**. The evolution of sexes. *Experientia Supplementum* **55**: 59–91.

Hoekstra RF, Janz RF, Schilstra AJ. **1984**. Evolution of gamete motility differences I. Relation between swimming speed and pheromonal attraction. *Journal of Theoretical Biology* **107**: 57–70.

Honda D, Shono T, Kimura K, Fujita S, Iseki M, Makino Y, Murakami A. **2007**. Homologs of the Sexually Induced Gene 1 (sig1) product constitute the Stramenopile mastigonemes. *Protist* **158**: 77–88.

Hori K, Matsubara K, Miyazawa K. **2000**. Primary structures of two hemagglutinins from the marine red alga, Hypnea japonica. *Biochimica et biophysica acta* **1474**: 226–236.

Hori K, Sato Y, Ito K, Fujiwara Y, Iwamoto Y, Makino H, Kawakubo A. **2007**. Strict specificity for high-mannose type N-glycans and primary structure of a red alga Eucheuma serra lectin. *Glycobiology* **17**: 479–491.

Hughes J. **1999**. Identification of a human homologue of the sea urchin receptor for egg jelly: a polycystic kidney disease-like protein. *Human Molecular Genetics* **8**: 543–549.

Hurst LD. **1990**. Parasite diversity and the evolution of diploidy, multicellularity and anisogamy. *Journal of theoretical biology* **144**: 429–443.

Hurst LD, Hamilton WD. **1992**. Cytoplasmic fusion and the nature of sexes. *Proceedings of the Royal Society B: Biological Sciences* **247**: 189–194.

Hurst LD, Nurse P. **1991**. A note on the evolution of meiosis. *Journal of Theoretical Biology* **150**: 561–563.

Hurst GDD, Werren JH. **2001**. The role of selfish genetic elements in eukaryotic evolution. *Nature Reviews Genetics* **2**: 597–606.

Hutson V, Law R. **1993**. Four steps to two sexes. *Proceedings of the Royal Society B: Biological Sciences* **253**: 43–51.

Janzen FJ, Krenz JG. **2005**. Phylogenetics: Which was first, TSD or GSD? In Temperature dependent sex determination in vertebrates. Temperature-dependent sex determination in vertebrates. Smithsonian Institution Scholarly Press, 121–130.

Jarne P. **1993**. The evolution of the selfing rate in functionally hermaphrodite plants and animals. *Annual Review of Ecology and Systematics* **24**: 441–466.

Johnstone RA, Keller L. **2000**. How males can gain by harming their mates: sexual conflict, seminal toxins, and the cost of mating. *The American Naturalist* **156**: 368–377.

Jokela J, Dybdahl MF, Lively CM. **2009**. The maintenance of sex, clonal dynamics, and host-parasite coevolution in a mixed population of sexual and asexual snails. *The American Naturalist* **174**: S43–S53.

Kalshoven H, Musgrave A, Vandenende H. **1990**. Mating receptor complex in the flagellar membrane of Chlamydomonas eugametos gametes. *Sexual Plant Reproduction* **3**: 77–87.

Kapranov P, Laurent GS, Raz T, Ozsolak F, Reynolds CP, Sorensen PH, Reaman G, Milos P, Arceci RJ, Thompson JF, *et al.* 2010. The majority of total nuclear-encoded non-ribosomal RNA in a human cell is 'dark matter' un-annotated RNA. *BMC Biology* **8**: 149.

Kawai H. **1992**. A summary of the morphology of chloroplasts and flagellated cells in the Phaeophyceae. *Korean Journal of Phycology* **7**: 33–43.

Kawai H, Müller DG, Fölster E, Häder D-P. **1990**. Phototactic responses in the gametes of the brown alga, Ectocarpus siliculosus. *Planta* **182**: 292–297.

Kelley LA, Sternberg MJE. **2009**. Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols* **4**: 363–371.

Kelly JK, Noor MAF. **1996**. Speciation by reinforcement: a model derived from studies of Drosophila. *Genetics* **143**: 1485–1497.

Kim GH, Fritz L. **1993**. A signal glycoprotein with α-d-mannosyl residues is involved in the wound-healing response of Antithamnion sparsum (Ceramiales, Rhodophyta). *Journal of Phycology* **29**: 85–90.

Kim S, Kim G. **1999**. Cell-cell recognition during fertilization in red alga, Aglaothamnion oosumiense (Ceramiaceae, Rodophyta). *Hydrobiologia* **398/399**: 81–87.

Kim GH, Klochkova TA, Yoon KS, Song YS, Lee KP. **2006**. Purification and characterization of a lectin, bryohealin, involved in the protoplast formation of a marine green alga Bryopsis plumosa (Chlorophyta). *Journal of Phycology* **42**: 86–95.

Kimura M. **1977**. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* **267**: 275–276.

Kimura M, Maruyama T. **1966**. The mutational load with epistatic gene interactions in fitness. *Genetics* **54**: 1337–1351.

Kirkpatrick M, Servedio MR. **1999**. The reinforcement of mating preferences on an island. *Genetics* **151**: 865–884.

Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER. **2013**. The Next-Generation Sequencing revolution and its impact on genomics. *Cell* **155**: 27–38.

Kondrashov AS. **1982**. Selection against harmful mutations in large sexual and asexual populations. *Genetics Research* **40**: 325–332.

Kondrashov AS. **1988**. Deleterious mutations and the evolution of sexual reproduction. *Nature* **336**: 435–440.

Korbie DJ, Mattick JS. **2008**. Touchdown PCR for increased specificity and sensitivity in PCR amplification. *Nature protocols* **3**: 1452–1456.

Korpelainen H. **1998**. Labile sex expression in plants. *Biological Reviews* **73**: 157–180.

Kosakovsky Pond SL, Frost SDW, Muse SV. **2005**. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**: 676–679.

Kreimer G, Kawai H, Müller DG, Melkonian M. **1991**. Reflective properties of the stigma in male gametes of Ectocarpus siliculosus (Phaeophyceae) studied by confocal laser scanning microscopy. *Journal of Phycology* **27**: 268–276.

Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL. **2001**. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journal of Molecular Biology* **305**: 567–580.

Kropf D. **1992**. Establishment and expression of cellular polarity in fucoid zygotes. *Microbiological Reviews* **56**: 316–339.

Kusano K, Sunohara Y, Takahashi N, Yoshikura H, Kobayashi I. **1994**. DNA double-strand break repair: genetic determinants of flanking crossing-over. *Proceedings of the National Academy of Sciences of the United States of America* **91**: 1173–1177.

Laemmli UK. **1970**. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680–685.

Lee N, Bakkeren G, Wong K, Sherwood JE, Kronstad JW. **1999**. The mating-type and pathogenicity locus of the fungus Ustilago hordei spans a 500-kb region. *Proceedings of the National Academy of Sciences* **96**: 15026–15031.

Lee YH, Ota T, Vacquier VD. **1995**. Positive selection is a general phenomenon in the evolution of abalone sperm lysin. *Molecular biology and evolution* **12**: 231–238.

Lehtonen J, Jennions MD, Kokko H. **2012**. The many costs of sex. *Trends in Ecology & Evolution* **27**: 172–178.

Lengeler KB, Fox DS, Fraser JA, Allen A, Forrester K, Dietrich FS, Heitman J. **2002**. Mating-type locus of Cryptococcus neoformans: a step in the evolution of sex chromosomes. *Eukaryotic Cell* **1**: 704–718.

Lessios HA, Lockhart S, Collin R, Sotil G, Sanchez-Jerez P, Zigler KS, Perez AF, Garrido MJ, Geyer LB, Bernardi G, *et al.* 2012. Phylogeography and bindin evolution in Arbacia, a sea urchin genus with an unusual distribution. *Molecular Ecology* **21**: 130–144.

Levin BR. **1988**. *The evolution of sex in bacteria. In The Evolution of sex: an examination of current ideas. ed. R. E. Michod, B. R. Levin, pp. 194- 211.* Sinauer Associates.

Levitan DR, Ferrell DL. **2006**. Selection on gamete recognition proteins depends on sex, density, and genotype frequency. *Science* **312**: 267–269.

Levitan DR, Fukami H, Jara J, Kline D, McGovern TM, McGhee KE, Swanson CA, Knowlton N. **2004**. Mechanisms of reproductive isolation among sympatric broadcast-spawning corals of the Montastraea annularis species complex. *Evolution; international journal of organic evolution* **58**: 308–323.

Lewin R. **1956**. Control of sexual activity in Chlamydomonas by light. *Journal of General Microbiology* **15**: 170–&.

Lewis WM. **1983**. Interruption of synthesis as a cost of sex in small organisms. *American Naturalist* **121**: 825–834.

Lewis WM Jr. **1987**. The cost of sex. The evolution of sex and its consequences (Stearns, S.C., ed.). Basel and Boston, .

Liu Y, Misamore MJ, Snell WJ. **2010**. Membrane fusion triggers rapid degradation of two gamete-specific, fusion-essential proteins in a membrane block to polygamy in Chlamydomonas. *Development (Cambridge, England)* **137**: 1473–1481.

Lobban CS, Wynne MJ. **1981**. *The Biology of Seaweeds*. University of California Press.

Loeser CR, Tulsiani DRP. **1999**. The role of carbohydrates in the induction of the acrosome reaction in mouse spermatozoa. *Biology of Reproduction* **60**: 94–101.

Lofthouse PF, Capon B. **1975**. Ultrastructural changes accompanying mitosporogenesis inEctocarpus parvus. *Protoplasma* **84**: 83–99.

Mah SA, Swanson WJ, Vacquier VD. **2005**. Positive selection in the carbohydrate recognition domains of sea urchin sperm receptor for egg jelly (suREJ) proteins. *Molecular Biology and Evolution* **22**: 533–541.

Maier I and S. **1995**. An immunofluorescence study on lectin binding sites in gametes of Ectocarpus siliculosus (Phaeophyceae). *Phycol. Res.* **43**: 33–42.

Maier I. **1997a**. The fine structure of the male gamete of Ectocarpus siliculosus (Ectocarpales, Phaeophyceae) .1. General structure of the cell. *European Journal of Phycology* **32**: 241–253.

Maier I. **1997b**. The fine structure of the male gamete of Ectocarpus siliculosus (Ectocarpales, Phaeophyceae) .2. The flagellar apparatus. *European Journal of Phycology* **32**: 255–266.

Mardis ER. **2008**. The impact of next-generation sequencing technology on genetics. *Trends in Genetics* **24**: 133–141.

Margulis DL, Sagan D. **1990**. *Origins of sex: Three billion years of genetic recombination (Bio-Origins Series)*. Yale University Press.

Martins MJF, Mota CF, Pearson GA. **2013**. Sex-biased gene expression in the brown alga Fucus vesiculosus. *BMC Genomics* **14**: 294.

Masters AK, Shirras AD, Hetherington AM. **1992**. Maternal mRNA and early development in Fucus serratus. *The Plant Journal* **2**: 619–622.

Matsuda H, Abrams PA. **1999**. Why are equally sized gametes so rare? The instability of isogamy and the cost of anisogamy. *Evolutionary Ecology Research* **1**: 769–784.

Maynard Smith J. **1978a**. *The Evolution of Sex*. Cambridge University Press.

Maynard Smith J. **1978b**. *The evolution of sex*. Cambridge University Press.

Maynard Smith J. **1990**. The evolution of prokaryotes: does sex matter? *Annual Review of Ecology and Systematics* **21**: 1–13.

Maynard Smith J, Dowson CG, Spratt BG. **1991**. Localized sex in bacteria. *Nature* **349**: 29–31.

Maynard Smith J, Szathmáry E. **1998**. *The major transitions in evolution*. Oxford University Press, USA.

Mayr E. **1942**. *Systematics and the origin of species, from the viewpoint of a zoologist*. Harvard University Press.

Mayr E (Biologe). **1964**. *Systematics and the origin of species*. Dover Publ.

McDaniel SF, Willis JH, Shaw AJ. **2007**. A Linkage map reveals a complex basis for segregation distortion in an interpopulation cross in the moss Ceratodon purpureus. *Genetics* **176**: 2489–2500.

McHugh DJ. **2003**. *A guide to the seaweed industry*. Food and Agriculture Organization of the United Nations.

Meagher TR. **2007**. Linking the evolution of gender variation to floral development. *Annals of Botany* **100**: 165–176.

Metz EC, Palumbi SR. **1996**. Positive selection and sequence rearrangements generate extensive polymorphism in the gamete recognition protein bindin. *Molecular Biology and Evolution* **13**: 397–406.

Metz EC, Robles-Sikisaka R, Vacquier VD. **1998**. Nonsynonymous substitution in abalone sperm fertilization genes exceeds substitution in introns and mitochondrial DNA. *Proceedings of the National Academy of Sciences* **95**: 10676–10681.

Michod RE, Levin BR. **1988**. *The Evolution of sex: an examination of current ideas*. Sinauer Associates.

Miller R, Wu G, Deshpande RR, Vieler A, Gärtner K, Li X, Moellering ER, Zäuner S, Cornish AJ, Liu B, *et al.* 2010. Changes in transcript abundance in Chlamydomonas reinhardtii following nitrogen deprivation predict diversion of metabolism. *Plant Physiology* **154**: 1737–1752.

Ming R, Moore PH. **2007**. Genomics of sex chromosomes. *Current Opinion in Plant Biology* **10**: 123–130.

Miyata T, Yasunaga T. **1980**. Molecular evolution of mRNA: A method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. *Journal of Molecular Evolution* **16**: 23–36.

Mori T, Barry R. ORC, Roberta Gardella SB, James B. McMahon MRB. **2005**. Isolation and characterization of griffithsin, a novel HIV-inactivating protein, from the red alga Griffithsia sp. *The Journal Of Biological Chemistry* **280**: 9345–9353.

Mori T, Kuroiwa H, Higashiyama T, Kuroiwa T. **2006**. Generative Cell Specific 1 is essential for angiosperm fertilization. *Nature cell biology* **8**: 64–71.

Moustafa A, Beszteri B, Maier UG, Bowler C, Valentin K, Bhattacharya D. **2009**. Genomic footprints of a cryptic plastid endosymbiosis in diatoms. *Science* **324**: 1724–1726.

Moy GW, Mendoza LM, Schulz JR, Swanson WJ, Glabe CG, Vacquier VD. **1996a**. The sea urchin sperm receptor for egg jelly is a modular protein with extensive homology to the human polycystic kidney disease protein, PKD1. *Journal of Cell Biology* **133**: 809–817.

Moy GW, Mendoza LM, Schulz JR, Swanson WJ, Glabe CG, Vacquier VD. **1996b**. The sea urchin sperm receptor for egg jelly is a modular protein with extensive homology to the human polycystic kidney disease protein, PKD1. *Journal of Cell Biology* **133**: 809–817.

Muller HJ. **1932**. *Some genetic aspects of sex*.

Muller HJ. **1964**. The relation of recombination to mutational advance. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **1**: 2–9.

Muller DG. **1976**. Sexual isolation between a european and american populations of Ectocarpus siliculosus (Pheophyta). *Journal of Phycology* **12**: 252–254.

Muller DG. **1979a**. Genetic affinity of Ectocarpus siliculosus (Dillw) Lyngb from the Mediterranean, North-Atlantic and Australia. *Phycologia* **18**: 312–318.

Muller DG. **1979b**. Olefinic hydrocarbons in sea-water: signal molecules for sexual reproduction in brown algae. *Pure and Applied Chemistry* **51**: 1885–1891.

Muller DG. **1988**. Studies on sexual compatibility between Ectocarpus siliculosus (Phaeophyceae) from Chile and the Mediterranean Sea. *Helgolander Meeresuntersuchungen* **42**: 469–476.

Müller DG. **1964**. Life-cycle of Ectocarpus siliculosus from Naples, Italy. *Nature* **203**: 1402–1402.

Müller DG. **1967**. Generationswechsel, kernphasenwechsel und sexualitat der braunalge Ectocarpus siliculosus im kulturversuch. *Planta* **75**: 39–54.

Müller DG. **1977**. Sexual reproduction in British Ectocarpus Siliculosus (Phaeophyta). *British Phycological Journal* **12**: 131–136.

Müller DG. **1979**. Genetic affinity of Ectocarpus siliculosus (Dillw.) Lyngb. from the Mediterranean, North Atlantic and Australia. *Phycologia* **18**: 312–318.

Müller DG. **1991**. Mendelian segregation of a virus genome during host meiosis in the marine brown alga Ectocarpus siliculosus. *Journal of Plant Physiology* **137**: 739–743.

Müller DG, Boland W, Jaenicke L, Gassmann G. **1985**. Diversification of chemoreceptors in Ectocarpus, Sphacelaria, and Adenocystis (Phaeophyceae). *Zeitschrift fur Naturforschung. Section C, Biosciences* **40**: 457–459.

Muller DG, Eichenberger W. **1995**. Crossing experiments, lipid composition, and the species concept in Ectocarpus siliculosus and E. fasciculatus (Pheophyceae, Ectocarpales). *Journal of Phycology* **31**: 173–176.

Müller DG, Falk H. **1973**. Flagellar structure of the gametes of Ectocarpus siliculosus (Phaeophyta) as revealed by negative staining. *Archiv für Mikrobiologie* **91**: 313–322.

Muller D, Jaenicke L, Donike M, Akintobi T. **1971**. Sex Attractant in a Brown Alga - Chemical Structure. *Science* **171**: 815–&.

Müller DG, Jaenicke L, Donike M, Akintobi T. **1971**. Sex attractant in a brown alga: chemical structure. *Science* **171**: 815–817.

Müller DG, Maier I, Müller H. **1987**. Flagellum autofluorescence and photoaccumulation in heterokont algae. *Photochemistry and Photobiology* **46**: 1003–1008.

Müller DG, Schmid CE. **1988**. Qualitative and quantitative determination of pheromone secretion in female gametes of Ectocarpus siliculosus (Pheophyceae). *Biological Chemistry Hoppe-Seyler* **396**: 647–653.

Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. **2012**. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet* **8**: e1002764.

Nagasato C, Motomura T. **2002**. Influence of the centrosome in cytokinesis of brown algae: polyspermic zygotes of Scytosiphon lomentaria (Scytosiphonales, Phaeophyceae). *Journal of Cell Science* **115**: 2541–2548.

Nedelcu AM, Marcu O, Michod RE. **2004**. Sex as a response to oxidative stress: a twofold increase in cellular reactive oxygen species activates sex genes. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **271**: 1591–1596.

Ning J, Otto TD, Pfander C, Schwach F, Brochet M, Bushell E, Goulding D, Sanders M, Lefebvre PA, Pei J, *et al.* 2013. Comparative genomics in Chlamydomonas and Plasmodium identifies an ancient nuclear envelope protein family essential for sexual reproduction in protists, fungi, plants, and vertebrates. *Genes & Development* **27**: 1198–1215.

Noor M. **1995**. Speciation driven by natural selection in Drosophila. *Nature* **375**: 674–675.

Noor MAF, Feder JL. **2006**. Speciation genetics: evolving approaches. *Nature Reviews Genetics* **7**: 851–861.

Nozawa M, Suzuki Y, Nei M. **2009**. Reliabilities of identifying positive selection by the branch-site and the site-prediction methods. *Proceedings of the National Academy of Sciences* **106**: 6700–6705.

O'Donnell L, Nicholls PK, O'Bryan MK, McLachlan RI, Stanton PG. **2011**. Spermiation. *Spermatogenesis* **1**: 14–35.

O'Keefe BR, Giomarelli B, Barnard DL, Shenoy SR, Chan PKS, McMahon JB, Palmer KE, Barnett BW, Meyerholz DK, Wohlford-Lenane CL, *et al.* 2010. Broad-spectrum in vitro activity and in vivo efficacy of the antiviral protein griffithsin against emerging viruses of the family Coronaviridae. *Journal of virology* **84**: 2511–2521.

Ortiz-Barrientos D, Counterman BA, Noor MAF. **2004**. The genetics of speciation by reinforcement. *PLoS Biol* **2**: e416.

Otto SP. **2009**. The evolutionary enigma of sex. *American Naturalist* **174**: S1–S14.

Palmer JD. **2003**. The symbiotic birth and spread of plastids: how many times and whodunit? *Journal of Phycology* **39**: 4–12.

Palumbi SR. **1994**. Genetic divergence, reproductive isolation, and marine speciation. *Annual Review of Ecology and Systematics* **25**: 547–572.

Palumbi SR. **1999**. All males are not created equal: Fertility differences depend on gamete recognition polymorphisms in sea urchins. *Proceedings of the National Academy of Sciences* **96**: 12632–12637.

Palumbi SR. **2008**. Speciation and the evolution of gamete recognition genes: pattern and process. *Heredity* **102**: 66–76.

Panhuis TM, Clark NL, Swanson WJ. **2006**. Rapid evolution of reproductive proteins in abalone and Drosophila. *Philosophical Transactions of the Royal Society B-Biological Sciences* **361**: 261–268.

Parker GA. **1979**. *Sexual selection and sexual conflict. In: Blum M.S, Blum N.A, editors. Sexual selection and reproductive competition in insects.* Academic Press.

Parker GA, Baker RR, Smith VGF. **1972**. The origin and evolution of gamete dimorphism and the male-female phenomenon. *Journal of Theoretical Biology* **36**: 529–553.

Parsch J, Ellegren H. **2013**. The evolutionary causes and consequences of sex-biased gene expression. *Nature Reviews Genetics* **14**: 83–87.

Pawley JB. **2010**. *Handbook of biological confocal microscopy*. Springer.

Pearson GA, Serrao EA. **2006**. Revisiting synchronous gamete release by fucoid algae in the intertidal zone: fertilization success and beyond? Annual Meeting of the Society-for-Integrative-and-Comparative-Biology. Orlando, FL, 587–597.

Peters AD, Lively CM. **1999**. The Red Queen and fluctuating epistasis: a population genetic analysis of antagonistic coevolution. *The American Naturalist* **154**: 393–405.

Peters AF, Mann AD, Cordova CA, Brodie J, Correa JA, Schroeder DC, Cock JM. **2010a**. Genetic diversity of Ectocarpus (Ectocarpales, Phaeophyceae) in Peru and northern Chile, the area of origin of the genome-sequenced strain. *New Phytologist* **188**: 30–41.

Peters AF, Scornet D, Müller DG, Kloareg B, Cock JM. **2004**. Inheritance of organelles in artificial hybrids of the isogamous multicellular chromist alga Ectocarpus siliculosus (Phaeophyceae). *European Journal of Phycology* **39**: 235–242.

Peters AF, Scornet D, Ratin M, Charrier B, Monnier A, Merrien Y, Corre E, Coelho SM, Cock M. **2008**. Life-cycle-generation-specific developmental processes are modified in the immediate upright mutant of the brown alga Ectocarpus siliculosus. *Development* **135**: 1503–1512.

Peters AF, Van Wijk SJ, Cho GY, Scornet D, Hanyuda T, Kawai H, Schroeder DC, Cock JM, Boo SM. **2010b**. Reinstatement of Ectocarpus crouaniorum Thuret in Le Jolis as a third common species of Ectocarpus (Ectocarpales, Phaeophyceae) in Western Europe, and its phenology at Roscoff, Brittany. *Phycological Research* **58**: 157–170.

Quatrano R. **1990**. *Model algal systems to study plant development* (H Marthy, Ed.). New York: Plenum Press Div Plenum Publishing Corp.

Rambaut A, Drummond A. **2009**. *Tracer v1.5, Available from http://beast.bio.ed.ac.uk/Tracer*.

Raper JR. **1966**. *Genetics of sexuality in higher fungi*. John Wiley & Sons, Incorporated.

Ravanko O. **1970**. *Morphological, developmental and taxonomic studies in the Ectocarpus complex (Phaeophyceae)*. Cramer.

Reyes-Prieto A, Weber APM, Bhattacharya D. **2007**. The origin and establishment of the plastid in algae and plants. *Annual Review of Genetics* **41**: 147–168.

Rice WR. **1987**. The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes. *Evolution* **41**: 911.

Rice WR, Holland B. **1997**. The enemies within: intergenomic conflict, interlocus contest evolution (ICE), and the intraspecific Red Queen. *Behavioral Ecology and Sociobiology* **41**: 1–10.

Rice DW, Palmer JD. **2006**. An exceptional horizontal gene transfer in plastids: gene replacement by a distant bacterial paralog and evidence that haptophyte and cryptophyte plastids are sisters. *BMC Biology* **4**: 31.

Robinson MD, Oshlack A. **2010**. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology* **11**: R25.

Ronquist F, Teslenko M, Van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. **2012**. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Systematic biology* **61**: 539–542.

Rumble SM, Lacroute P, Dalca AV, Fiume M, Sidow A, Brudno M. **2009**. SHRiMP: Accurate mapping of short color-space reads. *PLoS Computational Biology* **5**: e1000386.

Salathé M, Kouyos RD, Bonhoeffer S. **2009**. On the causes of selection for recombination underlying the Red Queen hypothesis. *The American Naturalist* **174**: S31–S42.

Salathé M, Kouyos RD, Regoes RR, Bonhoeffer S. **2008**. Rapid parasite adaptation drives selection for high recombination rates. *Evolution* **62**: 295–300.

Sanchez-Puerta MV, Delwiche CF. **2008**. A hypothesis for plastid evolution in Chromalveolates. *Journal of Phycology* **44**: 1097–1107.

Sanger F, Nicklen S, Coulson AR. **1977**. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences* **74**: 5463–5467.

Schmid CE. **1993**. Cell-cell-recognition during fertilization in Ectocarpus siliculosus (Phaeophyceae). *Hydrobiologia* **260/261**: 437–443.

Schmid CE, Muller DG, Eichenberger W. **1994a**. Isolation and characterization of a new phospholipid from brown algae intracellular localization and site of biosynthesis. *Journal of Plant Physiology* **143**: 570–574.

Schmid CE, Schroer N, Kawai H, Muller DG. **1992**. Isolation and biochemical-characterization of different gamete membranes in the chromophyte alga Ectocarpus siliculosus (Phaeophyceae). *Plant Physiology and Biochemistry* **30**: 703–712.

Schmid CE, Schroer N, Muller DG. **1994b**. Female gamete membrane glycoproteins potentially involved in gamete recognition in Ectocarpus siliculosus. *Plant Science* **102**: 61–67.

Servedio MR. **2000**. Reinforcement and the genetics of nonrandom mating. *Evolution* **54**: 21–29.

Silberfeld T, Leigh JW, Verbruggen H, Cruaud C, De Reviers B, Rousseau F. **2010**. A multi-locus time-calibrated phylogeny of the brown algae (Heterokonta, Ochrophyta, Phaeophyceae): Investigating the evolutionary nature of the 'brown algal crown radiation'. *Molecular Phylogenetics and Evolution* **56**: 659–674.

Simon SA, Zhai JX, Nandety RS, McCormick KP, Zeng J, Mejia D, Meyers BC. **2009**. Short-read sequencing technologies for transcriptional analyses. *Annu. Rev. Plant Biol.* **60**: 305–333.

Singh RS, Kulathinal RJ. **2000**. Sex gene pool evolution and speciation: A new paradigm. *Genes & Genetic Systems* **75**: 119–130.

Sonnhammer EL, Von Heijne G, Krogh A. **1998**. A hidden Markov model for predicting transmembrane helices in protein sequences. *Proceedings / ... International Conference on Intelligent Systems for Molecular Biology ; ISMB. International Conference on Intelligent Systems for Molecular Biology* **6**: 175–182.

Sorhannus U. **2003**. The effect of positive selection on a Sexual Reproduction Gene in Thalassiosira weissflogii (Bacillariophyta): results obtained from maximum-likelihood and parsimony-based methods. *Molecular Biology and Evolution* **20**: 1326–1328.

Sorhannus U, Kosakovsky Pond S. **2006**. Evidence for positive selection on a Sexual Reproduction Gene in the diatom genus Thalassiosira (Bacillariophyta). *Journal of Molecular Evolution* **63**: 231–239.

Spehr M, Schwane K, Riffell JA, Barbour J, Zimmer RK, Neuhaus EM, Hatt H. **2004**. Particulate adenylate cyclase plays a key role in human sperm olfactory receptor-mediated chemotaxis. *The Journal of Biological Chemistry* **279**: 40194–40203.

Springer SA, Moy GW, Friend DS, Swanson WJ, Vacquier VD. **2008**. Oyster sperm bindin is a combinatorial fucose lectin with remarkable intra-species diversity. *The International Journal of Developmental Biology* **52**: 759–768.

Stache B. **1990**. Sexual compatibility and species concept in Ectocarpus siliculosus (Ectocarpales, Pheophyceae) from Italy, North Carolina, Chile, and New Zealand. *Evolutionary biogeography of the marine algae of the North Atlantic* **G22**: 173–186.

Stache Crain B, Muller DG, Goff LJ. **1997**. Molecular systematics of Ectocarpus and Kuckuckia (Ectocarpales, Phaeophyceae) inferred from phylogenetic analysis of nuclear- and plastid-encoded DNA sequences. *Journal of Phycology* **33**: 152–168.

Stamatakis A. **2006**. Phylogenetic models of rate heterogeneity: a high performance computing perspective. Parallel and Distributed Processing Symposium, 2006. IPDPS 2006. 20th International.8 pp.–.

Stamatakis A, Hoover P, Rougemont J. **2008**. A rapid bootstrap algorithm for the RAxML Web servers. *Systematic biology* **57**: 758–771.

Sterck L, Billiau K, Abeel T, Rouzé P, Van de Peer Y. **2012**. ORCAE: online resource for community annotation of eukaryotes. *Nature Methods* **9**: 1041–1041.

Stratmann K, Boland W, Müller DG. **1993**. Biosynthesis of pheromones in female gametes of marine brown algae (Phaeophyceae). *Tetrahedron* **49**: 3755–3766.

Sugahara T, Ohama Y, Fukuda A, Hayashi M, Kawakubo A, Kato K. **2001**. The cytotoxic effect of Eucheuma serra agglutinin (ESA) on cancer cells and its application to molecular probe for drug delivery system using lipid vesicles. *Cytotechnology* **36**: 93–99.

Sutton KA, Jungnickel MK, Ward CJ, Harris PC, Florman HM. **2006**. Functional characterization of PKDREJ, a male germ cell-restricted polycystin. *Journal of cellular physiology* **209**: 493–500.

Swanson WJ. **1998**. Concerted evolution in an egg receptor for a rapidly evolving abalone sperm protein. *Science* **281**: 710–712.

Swanson WJ, Aagaard JE, Vacquier VD, Monné M, Sadat Al Hosseini H, Jovine L. **2011**. The molecular basis of sex: linking yeast to human. *Molecular Biology and Evolution* **28**: 1963 –1966.

Swanson WJ, Aquadro CF, Vacquier VD. **2001a**. Polymorphism in abalone fertilization proteins is consistent with the neutral evolution of the egg's receptor for lysin (VERL) and positive darwinian selection of sperm lysin. *Molecular Biology and Evolution* **18**: 376–383.

Swanson WJ, Vacquier VD. **1995a**. Extraordinary divergence and positive Darwinian selection in a fusagenic protein coating the acrosomal process of abalone spermatozoa. *Proceedings of the National Academy of Sciences* **92**: 4957–4961.

Swanson WJ, Vacquier VD. **1995b**. Liposome Fusion Induced by a Mr 18 000 Protein Localized to the Acrosomal Region of Acrosome-Reacted Abalone Spermatozoa. *Biochemistry* **34**: 14202–14208.

Swanson WJ, Vacquier VD. **1997**. The abalone egg vitelline envelope receptor for sperm lysin is a giant multivalent molecule. *Proceedings of the National Academy of Sciences* **94**: 6724–6729.

Swanson WJ, Vacquier VD. **1998**. Concerted evolution in an egg receptor for a rapidly evolving abalone sperm protein. *Science* **281**: 710–712.

Swanson WJ, Vacquier VD. **2002**. Reproductive protein evolution. *Annual Review of Ecology and Systematics* **33**: 161–179.

Swanson WJ, Wong A, Wolfner MF, Aquadro CF. **2004**. Evolutionary Expressed Sequence Tag analysis of Drosophila female reproductive tracts identifies genes subjected to positive selection. *Genetics* **168**: 1457–1465.

Swanson WJ, Yang Z, Wolfner MF, Aquadro CF. **2001b**. Positive darwinian selection drives the evolution of several female reproductive proteins in mammals. *Proceedings of the National Academy of Sciences* **98**: 2509–2514.

Szathmary E, Scheuring I, Kotsis M, Gladkih I. **1990**. Sexuality of eukaryotic unicells: hyperbolic growth, coexistence of facultative parthenogens, and the repair hypothesis. Organizational Constraints on the Dynamics of Evolution.(Maynard Smith, J. & Vida, G. eds). Manchester Univ. Press, .

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. **2011**. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular biology and evolution* **28**: 2731–2739.

Togashi T, Bartelt JL, Yoshimura J, Tainaka K, Cox PA. **2012**. Evolutionary trajectories explain the diversified evolution of isogamy and anisogamy in marine green algae. *Proceedings of the National Academy of Sciences* **109**: 13692–13697.

Trapnell C, Pachter L, Salzberg SL. **2009**. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**: 1105 –1111.

Treisman M, Dawkins R. **1976**. The 'Cost of Meiosis': is there any? *Journal of theoretical biology* **63**: 479–484.

Turner LM, Hoekstra HE. **2008**. Causes and consequences of the evolution of reproductive proteins. *The International Journal of Developmental Biology* **52**: 769–780.

Umen JG. **2011**. Evolution of sex and mating loci: An expanded view from Volvocine algae. *Current Opinion in Microbiology* **14**: 634–641.

Uyenoyama MK. **1984**. On the evolution of parthenogenesis: a genetic representation of the 'cost of meiosis'. *Evolution* **38**: 87.

Vacquier VD. **1998**. Evolution of gamete recognition proteins. *Science* **281**: 1995–1998.

Vacquier VD, Moy GW. **1997**. The fucose sulfate polymer of egg jelly binds to sperm REJ and is the inducer of the sea urchin sperm acrosome reaction. *Developmental Biology* **192**: 125–135.

Vanstechelman I. **2013**. Identification, characterization and evolution of mating type locus in diatoms.

Vanstechelman I, Sabbe K, Vyverman W, Vanormelingen P, Vuylsteke M. **2013**. Linkage mapping identifies the sex determining region as a single locus in the pennate diatom Seminavis robusta. *PLoS ONE* **8**: e60132.

Vilela-Silva A-CES, Hirohashi N, Mouro PAS. **2008**. The structure of sulfated polysaccharides ensures a carbohydrate-based mechanism for species recognition during sea urchin fertilization. *The International Journal of Developmental Biology* **52**: 551–559.

Villesen P. **2007**. FaBox: an online toolbox for fasta sequences. *Molecular Ecology Notes* **7**: 965–968.

Voelkerding KV, Dames SA, Durtschi JD. **2009**. Next-Generation Sequencing: from basic research to diagnostics. *Clinical Chemistry* **55**: 641–658.

Vrijenhoek RC. **1998**. Animal clones and diversity. *Bioscience* **48**: 617–628.

Wang Z, Gerstein M, Snyder M. **2009**. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10**: 57–63.

Ward CR, Storey BT, Kopf GS. **1992**. Activation of a G protein in mouse sperm membranes by solubilized proteins of the zona pellucida, the egg's extracellular matrix. *The Journal of Biological Chemistry* **267**: 14061–14067.

Weise L, Wiese W. **1978**. Sex cell contact in Chlamydomonas, a model for cell recognition. *Symposia of the Society for Experimental Biology* **32**: 83–103.

Weismann A, Poulton EB, Schönland S (Selmar), Shipley AE (Arthur E. **1889**. *Essays upon heredity and kindred biological problems*. Oxford : At the Clarendon Press.

Werner T. **2010**. Next Generation Sequencing in functional genomics. *Briefings in Bioinformatics* **11**: 499–511.

West JA, McBride DL. **1999**. Long-term and diurnal carpospore discharge patterns in the Ceramiaceae, Rhodomelaceae and Delesseriaceae (Rhodophyta). *Hydrobiologia* **398-399**: 101–114.

Wetterbom A, Ameur A, Feuk L, Gyllensten U, Cavelier L. **2010**. Identification of novel exons and transcribed regions by chimpanzee transcriptome sequencing. *Genome Biology* **11**: R78.

Wiese L. **1974**. Nature of sex specific glycoprotein agglutinins in Chlamydomonas. *Annals of the New York Academy of Sciences* **234**: 383–395.

Williams GC. **1975**. *Sex and evolution*. Princeton University Press.

Willie J Swanson, Vacquier D Victor. **2002**. The rapid evolution of reproductive proteins. *Nature reviews genetics* **3**.

Wong JL, Johnson MA. **2010**. Is HAP2-GCS1 an ancestral gamete fusogen? *Trends Cell Biol.* **20**: 134–141.

Wuest SE, Vijverberg K, Schmidt A, Weiss M, Gheyselinck J, Lohr M, Wellmer F, Rahnenfuhrer J, Von Mering C, Grossniklaus U. **2010**. Arabidopsis female gametophyte gene expression map reveals similarities between plant and animal gametes. *Current Biology* **20**: 506–512.

Yamagishi T, Motomura T, Nagasato C, Kawai H. **2009**. Novel proteins comprising the Stramenopile tripartite mastigoneme in Ochromonas danica (Chrysophyceae). *Journal of Phycology* **45**: 1110–1115.

Yamato KT, Ishizaki K, Fujisawa M, Okada S, Nakayama S, Fujishita M, Bando H, Yodoya K, Hayashi K, Bando T, *et al.* 2007. Gene organization of the liverwort Y chromosome reveals distinct sex chromosome evolution in a haploid system. *Proceedings of the National Academy of Sciences* **104**: 6472–6477.

Yang Z. **1998**. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Molecular biology and evolution* **15**: 568–573.

Yang Z. **2000**. Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus A. *Journal of molecular evolution* **51**: 423–432.

Yang Z. **2002**. Inference of selection from multiple species alignments. *Current opinion in genetics & development* **12**: 688–694.

Yang Z. **2007**. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution* **24**: 1586–1591.

Yang Z, Nielsen R, Goldman N, Pedersen A-MK. **2000a**. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* **155**: 431–449.

Yang Z, Swanson WJ, Vacquier VD. **2000b**. Maximum-likelihood analysis of molecular adaptation in abalone sperm lysin reveals variable selective pressures among lineages and sites. *Molecular Biology and Evolution* **17**: 1446–1455.

Zhang Z, Hambuch TM, Parsch J. **2004**. Molecular evolution of sex-biased genes in Drosophila. *Molecular Biology and Evolution* **21**: 2130–2139.

Zigler KS. **2008**. The evolution of sea urchin sperm bindin. *The International Journal of Developmental Biology* **52**: 791–796.